



99 43 39.1 255 6 ABP98904  
100 43 39.1 255 7 ADC31420

Abp98904 Human mol  
ADC31420 Human nov

## ALIGNMENTS

## RESULT 1

ADD93657

ADD93657 standard; protein; 1590 AA.

AC ADD93657;

DT 29-JAN-2004 (first entry)

DE Streptococcus sobrinus glucosyltransferase-I.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus sobrinus.

XX MO2003075845-A2.

XX 18-SEP-2003.

PF 07-MAR-2003; 2003MO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX Claim 16; Page 14; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diabetic or multi-epitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1590 AA;

XX Query Match 98.2%; Score 108; DB 7; Length 1590;

XX Best Local Similarity 95.5%; Pred. No. 3.6e-09; Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22

DB 548 VPSYSFARAHDSVQDLIRDI 569

RESULT 2  
AAR32925  
ID AAR32925 standard; protein; 1592 AA.

XX AAR32925;  
XX DT 28-JUN-1993 (first entry)

XX Glucosyltransferase I.  
DE GTF-I; Streptococcus; dental; caries.

XX Streptococcus sobrinus.

XX JP05023188-A.

XX 02-FEB-1993.

XX 25-JUL-1991; 91JP-00186592.

XX 25-JUL-1991; 91JP-00186592.

XX (KATO/) KATO K.

XX (FUKU/) FUKUI I.

XX WPI; 1993-079449/10.

XX N-PDB; AAQ37760.

XX DNA sequence glucosyltransferase-I - comprises Streptococcus sobrinus DNA sequence with at least one nucleotide added or deleted.

XX Claim 13; Page 15; 29pp; Japanese.

XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes glucosyltransferase-I (and mutants). The DNA was obtd. by treating S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109 transformed with it. A CC GTF-I expressing clone was isolated and sequenced. The clone may be used in the development of a drug for dental caries

XX Sequence 1592 AA;

XX Query Match 98.2%; Score 108; DB 2; Length 1592;

XX Best Local Similarity 95.5%; Pred. No. 3.6e-09; Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22

DB 548 VPSYSFARAHDSVQDLIRDI 569

RESULT 3  
AAU79285  
ID AAU79285 standard; protein; 1017 AA.

XX AAU79285;

XX 13-AUG-2002 (first entry)

XX Streptococcus mutans monoclonal antibody-related protein #2.

XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;

XX glucosyl transferase-B; immunotherapy.

XX Streptococcus mutans.

XX JP2002114709-A.

XX 16-APR-2002.

XX 04-OCT-2000; 2000JP-00304889.

XX 04-OCT-2000; 2000JP-00304889.

XX (UYNI-) UNIV NIPPON.

XX WPI; 2002-448885/48.

XX Anti-carries agent composed of a monoclonal antibody against an inhibitory

PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 XX  
 PS Claim 4; Page 17-19; 28pp; Japanese.  
 CC The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein  
 CC  
 SQ Sequence 1017 AA;  
 Query Match 95.5%; Score 105; DB 5; Length 1017;  
 Best Local Similarity 95.5%; Pred. No. 6.9e-09;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 VPSSYFARAHSEVQDLIRDI 22  
 |||||  
 Db 518 VPSSYFIRAHSEVQDLIRDI 539  
 |||||  
 RESULT 4  
 ID AAU79284 standard; protein; 1476 AA.  
 AC AAU79284;  
 XX  
 DT 13-AUG-2002 (first entry)  
 DE Streptococcus mutans monoclonal antibody-related protein #1.  
 XX  
 KW Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
 KW glucosyl transferase-B; immunotherapy.  
 XX  
 OS Streptococcus mutans.  
 PA JP2002114709-A.  
 PD 16-APR-2002.  
 XX  
 PF 04-OCT-2000; 2000JP-00304889.  
 XX  
 PR 04-OCT-2000; 2000JP-00304889.  
 XX  
 PA (UYN1-) UNITV NIPPON.  
 DR WPI; 2002-448885/48.  
 XX  
 PT Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 XX  
 PS Claim 3; Page 13-16; 28pp; Japanese.  
 CC The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein  
 CC  
 SQ Sequence 1476 AA;

Query Match 95.5%; Score 105; DB 5; Length 1476;  
 Best Local Similarity 95.5%; Pred. No. 1.1e-08;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 VPSSYFARAHSEVQDLIRDI 22  
 |||||  
 Db 552 VPSSYFIRAHSEVQDLIRDI 573  
 |||||  
 RESULT 5  
 ID AAU98028 standard; protein; 1375 AA.  
 AC AAU98028;  
 XX  
 DT 27-AUG-2002 (first entry)  
 DE S. mutans glucosyltransferase GTFC.  
 XX  
 KW Glucosyltransferase; GTFC; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture.  
 XX  
 OS Streptococcus mutans.  
 PA US2002031826-A1.  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007939.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 PI Nichols SR;  
 XX  
 DR WPI; 2002-414332/44.  
 DR N-PSDB; ABK52939.  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Disclosure; Page 30-33; 44pp; English.  
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448Y, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I144V/D457N/D567T/D571K/K719Q/K1014T, Y169A/Y170A/Y171A, and K719Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing





DT 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB mutant D457N.  
 DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; muteln.  
 XX  
 XX Streptococcus mutans.  
 OS Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 457  
 FT /note= "Wild-type Asp substituted by Asn"  
 PN US2002031826-A1.  
 XX  
 XX 14-MAR-2002.  
 PD  
 XX 19-DEC-2000; 2000US-00740274.  
 PF  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 PA (NICH/) NICHOLS S E.  
 XX  
 XX Nichols SE;  
 PI  
 XX WPI; 2002-414332/44.  
 DE  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 XX Claim 36; Page: 44pp; English.

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 XX Sequence 1475 AA;  
 SQ  
 Query Match 90.0%; Score 99; DB 5; Length 1475;  
 Best Local Similarity 90.9%; Pred. No. 1.2e-07;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 VPSYSPARHDSREVQDLRDI 22  
 Db 552 VPSYSPARHDSREVQDLRDI 573  
 RESULT 9  
 AAU98040  
 ID AAU98040 standard; protein, 1475 AA.  
 AC  
 AC AAU98040;  
 XX  
 XX 27-AUG-2002 (first entry)  
 DT  
 DE S. mutans glucosyltransferase GTFB mutant K779Q.  
 XX  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; muteln.  
 XX  
 XX Streptococcus mutans.  
 OS Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 779  
 FT /note= "Wild-type Lys substituted by Gln"  
 PN US2002031826-A1.  
 XX  
 XX 14-MAR-2002.  
 PD  
 XX 19-DEC-2000; 2000US-00740274.  
 PF  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 PA (NICH/) NICHOLS S E.  
 XX  
 XX Nichols SE;  
 PI  
 XX WPI; 2002-414332/44.  
 DE  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 XX Claim 36; Page: 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF)  
 B polypeptide having changes at position from 1448V, D457N, D567T,  
 K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,  
 I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 complementary polynucleotide, a ribonucleic acid sequence encoding the  
 GTF mutant, an expression cassette comprising the polynucleotide operably  
 linked to a promoter, a vector comprising the expression cassette, host  
 cell introduced with the vector, a transgenic plant comprising the  
 vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 coating composition comprising a glucan produced in a plant transformed  
 with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 thermoplastic molecule or their combinations and/or vacuole or a maize line  
 the glucan is produced in the amyloplast and/or vacuole or a maize line  
 deficient in starch biosynthesis, transformed with a gene encoding a  
 glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 comprising the glucan (paper sizing/coating agent). The vector is useful  
 for producing a glucan in a plant. The method comprises transforming a  
 plant cell with the vector, growing the plant cell under plant growing  
 conditions to produce a regenerated plant and inducing expression of the  
 polynucleotide for a time sufficient to produce the glucan in the  
 regenerated plant, where the vector contains a transit sequence from  
 ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and



XX S. mutans glucosyltransferase GTFB mutant I448V.  
 DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH Misc-difference 448 /note= "Wild-type Ile substituted by Val"  
 FT US2002031826-A1.  
 XX 14-MAR-2002.  
 PD 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 PI WPI; 2002-414332/44.  
 DR Glucosyltransferase B or D protein useful for producing a glucan useful  
 XX as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 PS The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 CC SQ Sequence 1475 AA;  
 XX  
 XX Query Match 90.0%; Score 99; DB 5; Length 1475;  
 XX Best Local Similarity 90.9%; Pred. No. 1.2e-07;  
 XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 XX Cy 1 VPSTSPRAHDSREVODLIRDII 22  
 XX Db 552 VPSTSPRAHDSREVODLIRDII 573  
 XX  
 XX RESULT 12  
 XX AAU98039  
 XX ID AAU98039 standard; protein; 1475 AA.  
 XX AC AAU98039;  
 XX XX 27-AUG-2002 (first entry)  
 XX DE S. mutans glucosyltransferase GTFB mutant YYY169-171AAA.  
 XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX OS Streptococcus mutans.  
 XX OS Synthetic.  
 XX Key Location/Qualifiers  
 XX FH Misc-difference 169.171  
 XX FT /note= "Wild-type Tyr-Tyr-Tyr substituted by Ala-Ala-Ala"  
 XX PT US2002031826-A1.  
 XX PN 14-MAR-2002.  
 XX PD 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 PI WPI; 2002-414332/44.  
 DR Glucosyltransferase B or D protein useful for producing a glucan useful  
 XX as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 PS The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably

linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein of potato or the vacuole of sugar beet. CC is produced in the amyloplast of potato and additions to modified starch. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Sequence 1475 AA:

Query Match 90.0%; Score 99; DB 5; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 1.2e-07;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 VPSTSFARADSEVQDLIRDI 22  
Db 552 VPSTSFARADSEVQDLIRDI 573

RESULT 13  
ID AAU98027 standard; protein; 1475 AA.

AC AAU98027;

DT 27-AUG-2002 (first entry)

DE 5. mutans glucosyltransferase GTFB.

KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
KM coating composition; glucan; starch; latex; thermoplastic molecule;  
KM amyloplast; vacuole; paper manufacture.

OS Streptococcus mutans.

XX US2002031826-A1.

PD 14-MAR-2002.

PF 19-DEC-2000; 2000US-00740274.

PR 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 16-JAN-1998; 98US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 20-JAN-1998; 98US-00008172.

PR 11-DEC-1998; 98US-00009620.

XX (NICH/) NICHOLS S E.

PI Nichols SE;  
XX WPI: 2002-414332/44.  
XX DR N-PsDB; ABK52938.  
PT Glucosyltransferase B or D protein useful for producing a glucan useful  
PT as substitutes for and additions to modified starch and latexes in paper  
PT manufacture, comprises mutations in specific positions.  
PS Disclosure; Page 21-25; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, or a 1448V/D457N/D567T/D571K/K719Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T588D, T589D, N471D, N471D/T589D, and N471D/T589D. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents GTFB

Sequence 1475 AA:

Query Match 90.0%; Score 99; DB 5; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 1.2e-07;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 VPSTSFARADSEVQDLIRDI 22  
Db 552 VPSTSFARADSEVQDLIRDI 573

RESULT 14

ID ADD93654 standard; protein; 1475 AA.

AC ADD93654;

DT 29-JAN-2004 (first entry)

DE Streptococcus mutans glucosyltransferase-B.

KM Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

OS Streptococcus mutans.

XX WO2003075845-A2.

```

XX 18-SEP-2003.
PD
XX
XX 07-MAR-2003; 2003WO-US006962.
PF
XX
XX 07-MAR-2002; 2002US-0363209P.
PR
XX 08-AUG-2002; 2002US-0402483P.
PR
XX (FORS-) FORSYTH INST.
PA
XX Smith DJ, Taubman MA;
PI
XX WPI; 2003-845091/78.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 16; Page 12-13; 49pp; English.
XX
XX The present sequence is the protein sequence of Streptococcus mutans
CC glucosyltransferase-B (GTF-B). Peptide fragments of GTF-B, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GGBP)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Depliptic or
CC multi-epitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of Gbpb can be used in passive immunisation.
XX
XX Sequence 1475 AA;
SQ
Query Match 90.0%; Score 99; DB 7; Length 1475;
Best Local Similarity 90.9%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 VPSYSPARAHDSVODLIRDI 22
Db 552 VPSYSPARAHDSVODLIRDI 573
RESULT 15
AAU98035
ID AAU98035 standard; protein; 1475 AA.
XX
XX AAU98035;
AC
XX
XX 27-AUG-2002 (first entry)
DT
XX
XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.
DE
XX
XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutain.
XX
XX Streptococcus mutans.
OS
XX Synthetic.
XX
XX Key Location/Qualifiers
FH MISC-difference 457 /note="Wild-type Asp substituted by Asn"
FT MISC-difference 571 /note="Wild-type Asp substituted by Lys"
FT
XX
XX US2002031826-A1.
XX
XX 14-MAR-2002.
XX
XX 19-DEC-2000; 2000US-00740274.
XX

```

```

PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
XX (NICH/) NICHOLS S E.
XX
XX Nichols SE;
PI
XX WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.
XX
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K1790/K1014T, Y169A/Y170A/Y171A, and K7790 or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant, a paper sizing and/or
CC vector, a seed or tuber from the transgenic plant, a paper sized and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilizing the glucan produced by GTF, which utilizes
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
XX Sequence 1475 AA;
SQ
Query Match 83.6%; Score 92; DB 5; Length 1475;
Best Local Similarity 86.4%; Pred. No. 1.9e-06;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 VPSYSPARAHDSVODLIRDI 22
Db 552 VPSYSPARAHDSVODLIRDI 573
RESULT 16
AAU98034
ID AAU98034 standard; protein; 1475 AA.

```

XX AC AAU98034;  
 XX DT 27-AUG-2002 (first entry)  
 XX DE S. mutans glucosyltransferase GTFB mutant D457N/D567T.  
 XX KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX KM amyloplast; vacuole; paper manufacture; mutant; mutein.  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX FH Key Location/Qualifiers  
 FT MISC-difference 457 /note= "Wild-type Asp substituted by Asn"  
 FT MISC-difference 567 /note= "Wild-type Asp substituted by Thr"  
 FT MISC-difference 567 /note= "Wild-type Asp substituted by Thr"  
 XX PN US2002031826-A1.  
 XX PD 14-MAR-2002.  
 XX PF 19-DEC-2000; 2000US-00740274.  
 XX PR 07-JUN-1995; 95US-00478704.  
 XX PR 07-JUN-1995; 95US-00482711.  
 XX PR 07-JUN-1995; 95US-00485243.  
 XX PR 16-JAN-1998; 98US-00007999.  
 XX PR 16-JAN-1998; 98US-00008172.  
 XX PR 20-JAN-1998; 98US-00009620.  
 XX PR 11-DEC-1998; 98US-00210361.  
 XX PA (NICH/) NICHOLS S E.  
 XX PI Nichols SE;  
 XX DR WPI; 2002-414332/44.  
 XX PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX PS Claim 36; Page: 44pp; English.  
 XX CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T, or a  
 CC KI014T, D457N/D567T, D457N/D571K, D567T/D571K/KI014T, or a  
 CC I448V/D457N/D567T/D571K/KI014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTF mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 XX  
 XX Query Match 83.6%; Score 92; DB 5; Length 1475;  
 XX Best Local Similarity 86.4%; Pred. No. 1.9e-06;  
 XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 VPSYSPRAHDSHVQDLINDII 22  
 DB 552 VPSYSPRAHDSHVQTLADI 573  
 RESULT 17  
 AAU98032  
 ID AAU98032 standard; protein; 1475 AA.  
 XX AC AAU98032;  
 XX DT 27-AUG-2002 (first entry)  
 XX DE S. mutans glucosyltransferase GTFB mutant D567T.  
 XX KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX KM amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX OS Streptococcus mutans.  
 XX OS Synthetic.  
 XX FH Key Location/Qualifiers  
 FT MISC-difference 567 /note= "Wild-type Asp substituted by Thr"  
 FT MISC-difference 567 /note= "Wild-type Asp substituted by Thr"  
 XX PN US2002031826-A1.  
 XX PD 14-MAR-2002.  
 XX PF 19-DEC-2000; 2000US-00740274.  
 XX PR 07-JUN-1995; 95US-00478704.  
 XX PR 07-JUN-1995; 95US-00482711.  
 XX PR 07-JUN-1995; 95US-00485243.  
 XX PR 16-JAN-1998; 98US-00007999.  
 XX PR 16-JAN-1998; 98US-00008172.  
 XX PR 20-JAN-1998; 98US-00009620.  
 XX PR 11-DEC-1998; 98US-00210361.  
 XX PA (NICH/) NICHOLS S E.  
 XX PI Nichols SE;  
 XX DR WPI; 2002-414332/44.  
 XX PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX PS Claim 36; Page: 44pp; English.  
 XX CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T, or a  
 CC KI014T, D457N/D567T, D457N/D571K, D567T/D571K/KI014T, or a  
 CC I448V/D457N/D567T/D571K/KI014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from 1589D, 1589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTF mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

CC SQ Sequence 1475 AA;

CC Query Match 83.6%; Score 92; DB 5; Length 1475;

CC Best Local Similarity 86.4%; Pred. No. 1.9e-06;

CC Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CC QY 1 VPSYSPARAHSEVQDLIRDI 22

DB 552 VPSYSPIRAHSEVQTLIADI 573

CC RESULT 18

CC AAU98036 ID AAU98036 standard; protein; 1475 AA.

CC AC AAU98036;

CC DT 27-AUG-2002 (first entry)

CC S. mutans glucosyltransferase GTFB mutant D567T/D571K.

CC KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;

CC KM coating composition; glucan; starch; latex; thermoplastic molecule;

CC KM amyloplast; vacuole; paper manufacture; mutant; muten.

CC OS Streptococcus mutans.

CC OS Synthetic.

CC XX Key Location/Qualifiers

CC FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"

CC FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

CC PN US2002031826-A1.

CC XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.  
 PF 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007998.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S R.

XX Nichols SE;

XX WPI; 2002-414332/44.

PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from 1589D, 1589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

CC SQ Sequence 1475 AA;

CC Query Match 77.3%; Score 85; DB 5; Length 1475;

CC Best Local Similarity 81.8%; Pred. No. 3e-05;

CC Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CC QY 1 VPSYSPARAHSEVQDLIRDI 22

DB 552 VPSYSPIRAHSEVQTLIADI 573



RESULT 19  
AAU98037  
ID AAU98037 standard; protein; 1475 AA.  
XX  
XX AAU98037;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.  
XX  
XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
XX coating composition; glucan; starch; latex; thermoplastic molecule;  
XX amyloplast; vacuole; paper manufacture; mutant; mutein.  
XX  
XX Streptococcus mutans.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"  
FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"  
FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"  
XX  
XX US2002031826-A1.  
XX  
XX 14-MAR-2002.  
XX  
XX 19-DEC-2000; 2000US-00740274.  
XX  
XX 07-JUN-1995; 95US-00478704.  
XX 07-JUN-1995; 95US-00482711.  
XX 07-JUN-1995; 95US-00485243.  
XX 16-JAN-1998; 98US-00007999.  
XX 16-JAN-1998; 98US-00008172.  
XX 20-JAN-1998; 98US-00009620.  
XX 11-DEC-1998; 98US-00210361.  
XX  
XX (NICH/) NICHOLS S E.  
XX  
XX Nichols SE;  
XX  
XX WPI; 2002-414332/44.  
XX  
XX Glucosyltransferase B or D protein useful for producing a glucan useful  
XX as substitutes for and additions to modified starch and latexes in paper  
XX manufacture, comprises mutations in specific positions.  
XX  
XX Claim 36; Page: 44pp; English.  
XX  
XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
XX B polypeptide having changes at position from 1448V, D457N, D567T,  
XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T, or a  
XX 1448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
XX GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
XX complementary polynucleotide, a ribonucleic acid sequence encoding the  
XX GTF mutant, an expression cassette comprising the polynucleotide operably  
XX linked to a promoter, a vector comprising the expression cassette, host  
XX cell introduced with the vector, a transgenic plant comprising the  
XX vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
XX coating composition comprising a glucan produced in a plant transformed  
XX with a gene encoding the mutant GTF, wild type or, starch, a latex,  
XX thermoplastic molecule or their combinations or glucan and starch where  
XX the glucan is produced in the amyloplast and/or vacuole or a maize line  
XX deficient in starch biosynthesis, transformed with a gene encoding a  
XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
XX comprising the glucan (paper sizing/coating agent). The vector is useful  
XX for producing a glucan in a plant. The method comprises transforming a  
XX plant cell with the vector, growing the plant cell under plant growing  
XX conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTF mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTFB sequence appearing as AAU98027 and the information in  
CC claim 36  
XX  
XX Sequence 1475 AA;  
XX  
XX  
XX Query Match 77.3%; Score 85; DB 5; Length 1475;  
XX Best Local Similarity 81.8%; Pred. No. 3e-05;  
XX Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
XX  
XX QY 1 VPSYSPRAHDSRVQDLIRDI 22  
XX Db 552 VPSYSPRAHDSRVQDLIRDI 573  
XX  
XX  
XX RESULT 20  
XX AAU98038  
XX ID AAU98038 standard; protein; 1475 AA.  
XX  
XX AAU98038;  
XX  
XX 27-AUG-2002 (first entry)  
XX  
XX  
XX S. mutans GTFB mutant I448V/D457N/D567T/D571K/K779Q/K1014T.  
XX  
XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
XX coating composition; glucan; starch; latex; thermoplastic molecule;  
XX amyloplast; vacuole; paper manufacture; mutant; mutein.  
XX  
XX Streptococcus mutans.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 448 /note= "Wild-type Ile substituted by Val"  
FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"  
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"  
FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"  
FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"  
FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"  
XX  
XX US2002031826-A1.  
XX  
XX 14-MAR-2002.  
XX  
XX 19-DEC-2000; 2000US-00740274.  
XX  
XX 07-JUN-1995; 95US-00478704.  
XX 07-JUN-1995; 95US-00482711.  
XX 07-JUN-1995; 95US-00485243.  
XX 16-JAN-1998; 98US-00007999.  
XX 16-JAN-1998; 98US-00008172.  
XX 20-JAN-1998; 98US-00009620.  
XX 11-DEC-1998; 98US-00210361.  
XX



PA (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

PT Glucosyltransferase B or D protein useful for producing a glucan useful  
PT as substitutes for and additions to modified starch and latexes in paper  
PT manufacture, comprises mutations in specific positions.

PS Claim 36; Page: 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
CC B polypeptide having changes at position from 1448V, D457N, D567T,  
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
CC GTF mutant, an expression cassette comprising the polynucleotide operably  
CC linked to a promoter, a vector comprising the expression cassette, host  
CC cell introduced with the vector, a transgenic plant comprising the  
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
CC coating composition comprising a glucan produced in a plant transformed  
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
CC thermoplastic molecule or their combinations or glucan and starch where  
CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTFB mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTFB sequence appearing as AAU98027 and the information in  
CC claim 36

XX Sequence 1475 AA;

XX Query Match 77.3%; Score 85; DB 5; Length 1475;

XX Best Local Similarity 81.8%; Pred No. 3e-05;

XX Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDLIRDI 22  
DB 552 VPSYSPARAHSEVQDLIRDI 573

RESULT 21

ID ABR63235 standard; protein; 2055 AA.

XX ABR63235;

DT 23-OCT-2003 (revised)

DT 27-AUG-2003 (first entry)

XX Glucanucrase sequence from strain KG15.

KW Glucan; glucosyltransferase activity; chickenery; prebiotic;  
KW bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.  
OS Lactobacillus sakei.

XX MO2003008618-A2.

XX 30-JAN-2003.

PF 22-JUL-2002; 2002MO-NL000495.

PR 20-JUL-2001; 2001EP-00202752.

PR 25-JUL-2001; 2001EP-00202841.

XX (NEDR ) NEDERLANDSE ORG TOEGEPAST.

PI Van Geel- Schutten GH;

XX WPI; 2003-289780/28.

DR N-PSDB; ACC84451.

PT Novel glucan produced by glucosyltransferase activity of lactic acid  
PT bacterium on sucrose substrate, and having backbone consisting of alpha  
PT (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.

XX Claim 11; Fig 1; 51pp; English.

XX The present invention relates to glucan capable of being produced by  
CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
CC substrate. The method is useful as a thickener, as a prebiotic and as a  
CC bioactive agent and as an anti-corrosion agent. The glucan can be  
CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
CC products. The glucan is useful as anticorrosion agent, e.g. for the  
CC protection of ship hulls. It can also be incorporated in nutritional or  
CC pharmaceutical compositions intended for improving the condition of the  
CC gastrointestinal tract. The present sequence represents a sequence of the  
CC glucanucrase gene. (Updated on 23-OCT-2003 to standardise OS field)

XX Sequence 2055 AA;

XX Query Match 76.4%; Score 84; DB 6; Length 2055;

XX Best Local Similarity 68.2%; Pred No. 6.5e-05;

XX Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDLIRDI 22  
DB 955 IENYSFVARAHSEVQTVIGDIV 976

RESULT 22

ID ABR63236 standard; protein; 1149 AA.

XX ABR63236;

DT 27-AUG-2003 (first entry)

XX Glucanucrase sequence from strain LB33.

XX Glucan; glucosyltransferase activity; thickener; prebiotic;

XX bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

XX Lactobacillus fermentum.

XX MO2003008618-A2.

XX 30-JAN-2003.

PF 22-JUL-2002; 2002MO-NL000495.

PR 20-JUL-2001; 2001EP-00202752.

PR 25-JUL-2001; 2001EP-00202841.

PA (NEDE ) NEDERLANDSE ORG TOEGEPAST.  
 XX Van Geel - Schutten GH;  
 XX WPI; 2003-289780/28.  
 DR N-PSDB; ACC84452.  
 XX Novel glucan produced by glucosyltransferase activity of lactic acid  
 PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3) - and alpha (1,6)-linked anhydroglucose units, useful as thickener.  
 XX  
 PS Claim 11; Fig 1; SIPP; English.  
 CC The present invention relates to glucan capable of being produced by  
 CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
 CC substrate. The method is useful as a thickener, as a prebiotic and as a  
 CC bioactive agent and as an anti-corrosion agent. The glucan can be  
 CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
 CC products. The glucan is useful as anticorrosion agent, e.g. for the  
 CC protection of ship hulls. It can also be incorporated in nutritional or  
 CC pharmaceutical compositions intended for improving the condition of the  
 CC gastrointestinal tract. The present sequence represents a sequence of the  
 CC glucanucrase gene  
 CC  
 SQ Sequence 1149 AA;  
 Query Match 73.6%; Score 81; DB 6; Length 1149;  
 Best Local Similarity 71.4%; Pred. No. 0.00011;  
 Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 OY 2 PSYSPARAHDSVQDLIRDI 22  
 DB 217 PSYSPARAHDSVQTVIAEIV 237  
 RESULT 23  
 ABR63228  
 ID ABR63228 standard; protein; 223 AA.  
 XX ABR63228;  
 AC  
 XX 27-AUG-2003 (first entry)  
 DT  
 XX  
 DE Glucanucrase sequence from strain LB86.  
 XX  
 XX Glucan; glucosyltransferase activity; thickener; prebiotic;  
 KM bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.  
 XX  
 OS Leuconostoc sp.  
 XX  
 FN WO2003008618-A2.  
 PD  
 XX 30-JAN-2003.  
 PF  
 XX 22-JUL-2002; 2002WO-NL000495.  
 XX  
 PR 20-JUL-2001; 2001EP-00262752.  
 PR 25-JUL-2001; 2001EP-00202841.  
 XX  
 PA (NEDE ) NEDERLANDSE ORG TOEGEPAST.  
 XX  
 PI Van Geel - Schutten GH;  
 XX WPI; 2003-289780/28.  
 DR N-PSDB; ACC50072.  
 DR  
 PT Novel glucan produced by glucosyltransferase activity of lactic acid  
 PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3) - and alpha (1,6)-linked anhydroglucose units, useful as thickener.  
 XX  
 PS Claim 11; Page 27; SIPP; English.  
 XX The present invention relates to glucan capable of being produced by

CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
 CC substrate. The method is useful as a thickener, as a prebiotic and as a  
 CC bioactive agent and as an anti-corrosion agent. The glucan can be  
 CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
 CC products. The glucan is useful as anticorrosion agent, e.g. for the  
 CC protection of ship hulls. It can also be incorporated in nutritional or  
 CC pharmaceutical compositions intended for improving the condition of the  
 CC gastrointestinal tract. The present sequence represents a sequence of the  
 CC glucanucrase gene  
 CC  
 SQ Sequence 223 AA;  
 Query Match 72.7%; Score 80; DB 6; Length 223;  
 Best Local Similarity 68.2%; Pred. No. 2.3e-05;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 VPSYSPARAHDSVQDLIRDI 22  
 DB 140 IPNYSFVARHDSVQTVIAQII 161  
 RESULT 24  
 ADC54806  
 ID ADC54806 standard; protein; 1499 AA.  
 XX ADC54806;  
 AC  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX  
 DE Protein Seq ID11 related to L mesenteroides dextran sucrose protein.  
 XX  
 XX dextran sucrose; active centre zone; glucan; polysaccharide; dextran;  
 KM D-glucose; starch; cellulose; glucan manufacture; transduction;  
 KM enzyme-reaction product.  
 XX  
 OS Unidentified.  
 XX  
 FN JP2003111590-A.  
 PD  
 XX 15-APR-2003.  
 PD  
 XX 03-OCT-2001; 2001JP-00307067.  
 PF  
 XX 03-OCT-2001; 2001JP-00307067.  
 PR  
 XX (DOKU-) DOKURITSU GYOSHI HOJIN SHOKUHIN SOGO KEN.  
 PA  
 XX WPI; 2003-735670/70.  
 DR P-PSDB; ADC54814.  
 DR  
 XX Novel modified dextran sucrose which exchanges one site of active center  
 PT zone of dextran sucrose for active center zone of different types of  
 PT dextran sucrose, useful for manufacturing glucan.  
 XX  
 PS Example 1; SEQ ID NO 11; 28pp; Japanese.  
 XX  
 CC This invention relates to a modified dextran sucrose (DS) exchanging one  
 CC site of the active centre zone of a dextran sucrose for the active centre  
 CC zone of different types of dextran sucrose. Glucan is a polysaccharide  
 CC (for example dextran) which uses D-glucoses, such as a starch and a  
 CC cellulose, as a structural unit. The modified enzyme of the invention is  
 CC useful in the manufacture of glucan. The selection of the active centre  
 CC area of DS which carries out transduction, enables changes in structure  
 CC and character of an enzyme-reaction product and their application to  
 CC various uses. The present sequence is that of a protein which is related  
 CC to the Leuconostoc mesenteroides dextran sucrose protein and which was  
 CC used during the exemplification of the invention.  
 CC  
 SQ Sequence 1499 AA;  
 Query Match 71.8%; Score 79; DB 7; Length 1499;  
 Best Local Similarity 71.4%; Pred. No. 0.00033;  
 Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;



PF 07-MAR-2003; 2003WO-US006962.  
 XX  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX  
 PA (FORS-) FORSYTH INST.  
 PI Smith DJ, Taubman MA;  
 DR WPI; 2003-845091/78.  
 XX  
 PT Composition useful as vaccines for dental caries comprises a fragment of  
 a glucan binding protein-B binding to a major histocompatibility complex  
 class II protein.  
 XX  
 PS Claim 16; Page 15; 49pp; English.  
 CC The present sequence is the protein sequence of Streptococcus sobrinus  
 CC glucosyltransferase-U (GTF-U). Peptide fragments of GTF-U, especially  
 CC from the catalytic domain of the polypeptide, can be used in immunogenic  
 CC compositions and subunit vaccines for dental caries. These compositions  
 CC comprise a major histocompatibility complex (MHC) class II protein-  
 CC binding peptide from S. mutants glucan binding protein-B (GbpB)  
 CC covalently linked with a peptide fragment of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Diepitopic or  
 CC multipitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX  
 SQ Sequence 1554 AA;  
 XX  
 Query Match 70.9%; Score 78; DB 7; Length 1554;  
 Best Local Similarity 63.6%; Pred. No. 0.00051;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VPSYSPARAHDSVQDLIRDI 22  
 Db 557 IPTYSFVARHDSVQTVIAKIV 578  
 XX  
 RESULT 28  
 ABR63229  
 ID ABR63229 standard; protein; 221 AA.  
 XX  
 AC ABR63229;  
 XX  
 DT 27-AUG-2003 (first entry)  
 XX  
 DE Glucanase sequence from strain Lb86.  
 XX  
 KW Glucan; glucosyltransferase activity; chicken; prebiotic;  
 KW bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.  
 XX  
 OS Leuconostoc sp.  
 XX  
 PN WO2003008618-A2.  
 XX  
 PD 30-JAN-2003.  
 XX  
 PF 22-JUL-2002; 2002WO-NL000495.  
 XX  
 PR 20-JUL-2001; 2001EP-00202752.  
 PR 25-JUL-2001; 2001EP-00202841.  
 XX  
 PA (NEDE) NEDERLANDSE ORG TOEGEPAST.  
 PI Van Geel- Schutten GH;  
 DR WPI; 2003-289780/28.  
 DR N-PSDB; ACC50073.  
 XX  
 PT Novel glucan produced by glucosyltransferase activity of lactic acid

PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.  
 XX  
 PS Claim 11; Page 29; 51pp; English.  
 XX  
 CC The present invention relates to glucan capable of being produced by  
 CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
 CC substrate. The method is useful as a thickener, as a prebiotic and as a  
 CC bioactive agent and as an anti-corrosion agent. The glucan can be  
 CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
 CC products. The glucan is useful as anticorrosion agent, e.g. for the  
 CC protection of ship hulls. It can also be incorporated in nutritional or  
 CC pharmaceutical compositions intended for improving the condition of the  
 CC gastrointestinal tract. The present sequence represents a sequence of the  
 CC glucanase gene  
 XX  
 SQ Sequence 221 AA;  
 XX  
 Query Match 70.0%; Score 77; DB 6; Length 221;  
 Best Local Similarity 63.6%; Pred. No. 7.6e-05;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VPSYSPARAHDSVQDLIRDI 22  
 Db 139 IPNYSFVARHDSVQGIIGKIL 160  
 XX  
 RESULT 29  
 ABB98651  
 ID ABB98651 standard; peptide; 15 AA.  
 XX  
 AC ABB98651;  
 XX  
 DT 29-AUG-2003 (revised)  
 DT 14-JAN-2003 (first entry)  
 XX  
 DE Dextrane-saccharase, DSR-E, catalytic domain conserved peptide GTFI #5.  
 XX  
 KW Dextran saccharase; enzyme; cytotactic; dermatological; antiseborrheic;  
 KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;  
 KW microflora regulation; intestinal transit; mineral assimilation;  
 KW colon cancer; acne; dandruff; body odour.  
 XX  
 OS Leuconostoc mesenteroides; NRRL B-1299.  
 XX  
 PN FR2822163-A1.  
 XX  
 PD 20-SEP-2002.  
 XX  
 PF 19-DEC-2001; 2001FR-00016495.  
 XX  
 PR 16-MAR-2001; 2001FR-00003631.  
 XX  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 PI Bozonnet SAM, Remaud SMAC, Willemot RML, Monsan PEF;  
 DR WPI; 2002-715213/78.  
 XX  
 PT New glycosyl transferase enzymes, containing glucan bonding and catalytic  
 PT domains and producing alpha-(1-2) branched dextrans, useful in probiotic,  
 PT pharmaceutical or cosmetic compositions.  
 XX  
 PS Example 2; Fig 5; 82pp; French.  
 XX  
 CC The present invention relates to a novel dextran saccharase, DSR-E  
 CC (ABB98651), from Leuconostoc mesenteroides NRRL B-1299. The saccharase  
 CC has glycosyl transferase activity suitable for producing dextrans having  
 CC alpha(1-2) branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside,  
 CC alpha-fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbidopyranoside or  
 CC alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in  
 CC prebiotic, pharmaceutical or cosmetic compositions. The dextrans and  
 CC related compounds having alpha(1-2) bonds, produced using DSR-E, may be

CC involved in signalling/cellular recognition processes in vivo  
 CC (specifically in regulation of microflora in the intestines or on the  
 CC skin); and are potentially useful for improving intestinal transit,  
 CC increasing assimilation of minerals (e.g. calcium and/or magnesium),  
 CC preventing cancer of the colon and combating skin problems such as acne,  
 CC dandruff and body odour. The present sequence is a conserved peptide  
 CC sequence from the catalytic domain of the DSR-E of the invention.  
 CC (updated on 29-AUG-2003 to standardise OS field)

XX  
 SQ Sequence 15 AA;

Query Match 69.1%; Score 76; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 YSFARAHDSFVQDLI 18  
 |||||  
 Db 1 YSFARAHDSFVQDLI 15

RESULT 30

ABR63226  
 ID ABR63226 standard; protein; 221 AA.

XX  
 AC ABR63226;

XX  
 DT 27-AUG-2003 (first entry)

XX  
 DE Glucanucrase sequence from strain Lb180.

XX  
 KM Glucan; glucosyltransferase activity; thickener; prebiotic;

XX  
 KM bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

XX  
 OS Lactobacillus reuteri.

XX  
 PN WO2003008618-A2.

XX  
 PD 30-JAN-2003.

XX  
 PF 22-JUL-2002; 2002WO-NL000495.

XX  
 PR 20-JUL-2001; 2001EP-00202752.

XX  
 PR 25-JUL-2001; 2001EP-00202841.

XX  
 PA (NEDE ) NEDERLANDSE ORG TOEGEPAST.

XX  
 PI Van Geel - Schutten GH;

XX  
 DR WPI; 2003-289780/28.

XX  
 DR N-PSDB; ACC50070.

XX  
 PT Novel glucan produced by glucosyltransferase activity of lactic acid  
 PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.

XX  
 PS Claim 11; Page 24-25; 51pp; English.

XX  
 CC The present invention relates to glucan capable of being produced by  
 CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
 CC substrate. The method is useful as a thickener, as a prebiotic and as a  
 CC bioactive agent and as an anti-corrosion agent. The glucan can be  
 CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
 CC products. The glucan is useful as anticorrosion agent, e.g. for the  
 CC protection of ship hulls. It can also be incorporated in nutritional or  
 CC pharmaceutical compositions intended for improving the condition of the  
 CC gastrointestinal tract. The present sequence represents a sequence of the  
 CC glucanucrase gene

XX  
 SQ Sequence 221 AA;

Query Match 69.1%; Score 76; DB 6; Length 221;  
 Best Local Similarity 63.6%; Pred. No. 0.00011;  
 Matches 14; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSYFARAHDSFVQDLIRDI 22  
 :|||:|||||  
 Db 140 IPSYFARAHDSNAODQIRQAI 161

Search completed: May 4, 2004, 09:10:01  
 Job time : 50.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:06:31 ; Search time 11.6667 Seconds

(without alignments)  
181.390 Million cell updates/sec

Title: US-09-290-049a-17

Perfect score: 110

Sequence: 1 VPSYFARAHDSYQDLIRDI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: PIR\_78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	98.2	1592	2 A38175	glucosyltransferase
2	100	90.9	1375	2 JTO345	dextranase (EC
3	99	90.0	1475	2 B33135	glucosyltransferase
4	83	75.5	1290	2 UC5473	dextranase (EC
5	80	72.7	1508	2 T31098	glucosyltransferase
6	70	63.6	1365	2 A41483	glucosyltransferase
7	70	63.6	1599	2 S23737	glucosyltransferase
8	68	61.8	1449	2 T30857	glucosyltransferase
9	68	61.8	1449	2 T30857	glucosyltransferase
10	65	59.1	1518	2 A44811	glucosyltransferase
11	62	56.4	1431	2 A45866	glucosyltransferase
12	62	56.4	1577	2 T30858	glucosyltransferase
13	49	44.5	108	2 S75540	hypothetical prote
14	47	42.7	51	2 G82455	galactose-1-phosph
15	47	42.7	508	2 B83788	DNA polymerase III
16	47	42.7	597	2 F82935	exonuclease ABC C
17	46.5	42.3	603	2 AB1229	retrovirus-related
18	46.5	42.3	603	2 AD1582	atrazine chlorohyd
19	45	40.9	1058	1 GNFP17	hypothetical prote
20	44	40.0	330	2 F69471	hypothetical prote
21	44	40.0	506	2 T47184	hypothetical prote
22	44	40.0	540	1 OYHUCR	hypothetical prote
23	44	40.0	651	2 T16450	hypothetical prote
24	44	40.0	6359	2 T31679	hypothetical prote
25	43.5	39.5	484	2 AB8920	glutamy-tRNA synth
26	43	39.1	176	2 F83819	hypothetical prote
27	43	39.1	195	2 F72718	hypothetical prote
28	43	39.1	508	2 T50180	nucleolar protein
29	43	39.1	570	2 F70332	proline-tRNA synth

30	43	39.1	631	2 H90599	conserved hypothet
31	43	39.1	637	2 E75044	DNA repair helicase
32	42.5	38.6	245	2 AG1524	riboflavin kinase
33	42.5	38.6	246	2 AH1165	riboflavin kinase
34	42.5	38.6	486	2 F98106	glutamate-tRNA lig
35	42.5	38.6	540	2 S63299	sugar transport pr
36	42	38.2	188	2 S31626	cell fusion protei
37	42	38.2	196	2 S48302	MAD2 protein - yea
38	42	38.2	216	2 AF1336	L-fuculose-phospha
39	42	38.2	284	2 C64158	hypothetical prote
40	42	38.2	284	2 D96722	hypothetical prote
41	42	38.2	359	2 G82197	hypothetical prote
42	42	38.2	429	2 S20050	transcription fact
43	42	38.2	451	2 T41129	hypothetical prote
44	42	38.2	575	2 S46329	intermediate filam
45	42	38.2	635	2 T49694	hypothetical prote
46	42	38.2	822	2 F69164	ATP-dependent RNA
47	42	38.2	907	2 B66636	hypothetical prote
48	42	38.2	1146	2 E70204	hypothetical prote
49	42	38.2	1389	2 I58157	peritaxin - rat
50	42	38.2	1835	2 S46082	urea carboxylase (
51	41.5	37.7	941	2 AB1966	two-component sens
52	41	37.3	27	2 S55234	flavodoxin isoform
53	41	37.3	68	2 AC1093	hypothetical prote
54	41	37.3	139	2 T18916	hypothetical prote
55	41	37.3	295	1 A64236	hypothetical prote
56	41	37.3	361	2 A87656	hypothetical prote
57	41	37.3	419	2 A40728	microphthalmia-ass
58	41	37.3	444	2 F66329	hypothetical prote
59	41	37.3	475	2 AB2767	magnesium transpor
60	41	37.3	501	2 B97547	magnesium transpor
61	41	37.3	548	1 UFECAQ	funarate hydratase
62	41	37.3	548	2 F90918	funarase A [import
63	41	37.3	548	2 C85767	funarase A [isomila
64	41	37.3	548	2 AB0691	funarate hydratase
65	41	37.3	554	2 P86244	hypothetical prote
66	41	37.3	571	2 T08930	hypothetical prote
67	41	37.3	734	2 AB2047	hypothetical prote
68	41	37.3	933	1 DECCOG	primosomal replica
69	41	37.3	933	2 F85573	oxoglutarate dehyd
70	41	37.3	933	2 G90722	hypothetical prote
71	41	37.3	933	2 AD0591	oxoglutarate dehyd
72	41	37.3	953	2 S13520	2-oxoglutarate deh
73	41	37.3	953	2 T46913	coat protein beta
74	41	37.3	1040	2 S34639	hypothetical prote
75	40.5	36.8	196	2 T29046	pol protein - frui
76	40.5	36.8	240	2 A64652	hypothetical prote
77	40.5	36.8	240	2 D71939	hypothetical prote
78	40.5	36.8	270	2 F84255	hypothetical prote
79	40.5	36.8	395	2 B75047	hypothetical prote
80	40.5	36.8	484	2 S73490	glutamate-tRNA lig
81	40.5	36.8	688	2 S55349	glutamate-tRNA lig
82	40.5	36.8	767	2 T39715	potassium channel
83	40.5	36.8	956	2 H81654	probable transcrip
84	40.5	36.8	80	2 D78808	conserved hypothet
85	40	36.4	80	2 B69157	sp68.1 protein - M
86	40	36.4	179	2 A12056	hypothetical prote
87	40	36.4	205	2 C75155	polypeptide deform
88	40	36.4	207	2 AG0493	methylyransferase
89	40	36.4	217	2 T11382	superoxide dismuta
90	40	36.4	232	2 T03678	cytochrome-c oxida
91	40	36.4	267	2 A13010	plasma membrane pr
92	40	36.4	267	2 B98273	hypothetical prote
93	40	36.4	314	1 ITVZKA	hypothetical prote
94	40	36.4	327	2 H55094	ion (III) diclitrat
95	40	36.4	397	2 S26731	poxvirus DNA topol
96	40	36.4	397	2 S26731	hypothetical prote
97	40	36.4	415	2 T44601	neuro-14 protein -
98	40	36.4	415	2 A65936	probable glycosylt
99	40	36.4	419	2 C95376	probable glycosylt
100	40	36.4	429	2 B83826	probable arginine

## ALIGNMENTS

## RESULT 1

A38175  
 glucosyltransferase precursor - Streptococcus sobrinus  
 C/Species: Streptococcus sobrinus  
 C/Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
 C/Accession: A38175  
 R/AbO, H.; Kusunuma, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
 J. Bacteriol. 173, 989-996, 1991  
 A/Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
 A/Reference number: A38175; PMID:91123227; PMID:1704006  
 A/Accession: A38175  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1592 <ABO>  
 A/Cross-references: GB:D90213; NID:G217032; PIDN:BA14241.1; PID:dl014946; PID:g217033  
 C/Superfamily: cpl repeat homology  
 F:1093-1112/Domain: cpl repeat homology <CP1>  
 F:1222-1241/Domain: cpl repeat homology <CP2>  
 F:1287-1306/Domain: cpl repeat homology <CP3>  
 F:1330-1351/Domain: cpl repeat homology <CP4>  
 F:1352-1371/Domain: cpl repeat homology <CP5>  
 F:1402-1420/Domain: cpl repeat homology <CP6>  
 F:1465-1484/Domain: cpl repeat homology <CP7>  
 F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 98.2%; Score 108; DB 2; Length 1592;  
 Best Local Similarity 95.5%; Pred. No. 1.5e-08;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSPFARADSEVODLIIRDI 22  
 |||||  
 DB 548 VPSPFARADSEVODLIIRDI 569

## RESULT 2

JT0345  
 dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
 N/Alternate names: Sucrose 6-glucosyltransferase  
 C/Species: Streptococcus mutans  
 C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
 C/Accession: JT0345; C33135  
 R/Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
 Gene 69, 101-109, 1988  
 A/Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
 A/Reference number: JT0345; MUID:89137980; PMID:2976010  
 A/Accession: JT0345  
 A/Molecule type: DNA  
 A/Residues: 1-1375 <UDB>  
 A/Experimental source: GS-5  
 R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 J. Bacteriol. 169, 4263-4270, 1987  
 A/Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
 A/Reference number: A33135; MUID:87308013; PMID:3040685  
 A/Accession: C33135  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-349 <SHI>  
 A/Cross-references: GB:ML7361  
 C/Genetics:  
 C/Function: A/Gene: gtfC  
 A/Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
 C/Superfamily: cpl repeat homology  
 C/Keywords: duplication; glycosyltransferase; hexosyltransferase  
 F:1-34/Domain: signal sequence #status predicted <SIG>  
 F:35-1375/Product: glucosyltransferase #status predicted <MAT>  
 F:1126-1145/Domain: cpl repeat homology <CP1>  
 F:1253-1272/Domain: cpl repeat homology <CP2>  
 F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 90.9%; Score 100; DB 2; Length 1375;  
 Best Local Similarity 90.9%; Pred. No. 2.3e-07;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSPFARADSEVODLIIRDI 22  
 |||||  
 DB 578 VPSPFARADSEVODLIIRDI 599

## RESULT 3

B31135  
 gtfB protein precursor - Streptococcus mutans  
 C/Species: Streptococcus mutans  
 C/Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999  
 C/Accession: B31135; A33128  
 R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 J. Bacteriol. 169, 4263-4270, 1987  
 A/Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
 A/Reference number: A33135; MUID:87308013; PMID:3040685  
 A/Accession: B31135  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1475 <SHI>  
 A/Cross-references: GB:ML7361; NID:G153639; PIDN:AA88588.1; PID:G153640  
 R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 Submitted to the Protein Sequence Database, September 1990  
 A/Reference number: A33128  
 A/Accession: A33128  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-171,173-641,'N',643-1475 <SH2>  
 A/Experimental source: strain GS-5  
 C/Superfamily: cpl repeat homology  
 F:1096-1115/Domain: cpl repeat homology <CP1>  
 F:1224-1243/Domain: cpl repeat homology <CP2>  
 F:1289-1308/Domain: cpl repeat homology <CP3>  
 F:1354-1373/Domain: cpl repeat homology <CP4>  
 F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 90.0%; Score 99; DB 2; Length 1475;  
 Best Local Similarity 90.9%; Pred. No. 3.6e-07;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSPFARADSEVODLIIRDI 22  
 |||||  
 DB 552 VPSPFARADSEVODLIIRDI 573

## RESULT 4

JC5473  
 dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
 C/Species: Leuconostoc mesenteroides  
 C/Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
 C/Accession: JC5473  
 R/Monchois, V.; Willems, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.  
 Gene 182, 23-32, 1996  
 A/Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc  
 A/Reference number: JC5473; MUID:97136686; PMID:8962063  
 A/Accession: JC5473  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-1290 <MON>  
 A/Cross-references: GB:U38181  
 C/Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto  
 C/Genetics:  
 A/Gene: dsrA  
 C/Keywords: glycosyltransferase; hexosyltransferase  
 F:78-87/Domain: catalytic #status predicted <CAT>  
 F:922-1190/Domain: glucan-binding #status predicted <GCB>

Query Match 75.5%; Score 83; DB 2; Length 1290;  
 Best Local Similarity 76.2%; Pred. No. 0.00011;  
 Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PSYSPARAHSEVQDLIRDI 22  
 DB 388 PNYSFVRAHSEVQTVIAQII 408

## RESULT 5

T31098  
 Probable dextranucrase (EC 2.4.1.5), extracellular - leuconostoc mesenteroides  
 C:Species: Leuconostoc mesenteroides  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T31098  
 R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
 R:EMS Microbiol. Lett. 159, 307-315, 1998  
 A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (Dg  
 A:Reference number: Z20981; MUID:98164374; PMID:9503626  
 A:Accession: T31098  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1508 <MON>  
 A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AA895453.1  
 A:Experimental source: strain NRRL B-1299  
 C:Genetics:  
 A:Gene: dextr  
 C:Function:  
 A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 72.7%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 68.2%; Pred. No. 0.00039;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDLIRDI 22  
 DB 634 IPNYSFVRAHSEVQTVIAQII 655

## RESULT 6

A41483  
 glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
 C:Species: Streptococcus sobrinus  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1995  
 C:Accession: A41483  
 R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
 Infect. Immun. 58, 2452-2458, 1990  
 A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltrans  
 A:Reference number: A41483; MUID:90316665; PMID:2142479  
 A:Accession: A41483  
 A:Molecule type: DNA  
 A:Residues: 1-1365 <GIL>  
 A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653  
 C:Genetics:  
 A:Gene: gtfS  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 63.6%; Score 70; DB 2; Length 1365;  
 Best Local Similarity 68.2%; Pred. No. 0.013;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDLIRDI 22  
 DB 537 VPNYVFIKRAHSEVQTVIAKII 558

## RESULT 7

S22737  
 glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
 C:Species: Streptococcus salivarius  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
 C:Accession: S22737; S28810; B44811; S22727  
 R:Jacques, N.  
 submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737  
 A:Molecule type: DNA  
 A:Residues: 1-1599 <JAC>  
 A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
 A:Experimental source: ATCC 25975  
 R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
 J. Gen. Microbiol. 137, 2577-2593, 1991  
 A:Title: Molecular characterization of a cluster of at least two glucosyltransferase gen  
 A:Reference number: A44811; MUID:92148377; PMID:1838391  
 A:Accession: S28810  
 A:Molecule type: DNA  
 A:Residues: 1-51 <GIF>  
 A:Cross-references: EMBL:Z11873  
 C:Genetics:  
 A:Gene: gtfK  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 63.6%; Score 70; DB 2; Length 1599;  
 Best Local Similarity 70.0%; Pred. No. 0.016;  
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPARAHSEVQDLIRDI 22  
 DB 574 TYLAFVRAHSEVQTVIAQII 593

## RESULT 8

T30857  
 glucosyltransferase - Streptococcus salivarius  
 C:Species: Streptococcus salivarius  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30857  
 R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
 Infect. Immun. 63, 609-621, 1995  
 A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri  
 A:Reference number: Z20909; MUID:95122197; PMID:7822030  
 A:Accession: T30857  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1449 <SIM>  
 A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AA641412.1  
 C:Genetics:  
 A:Gene: gtfL

Query Match 61.8%; Score 68; DB 2; Length 1449;  
 Best Local Similarity 65.0%; Pred. No. 0.03;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPARAHSEVQDLIRDI 22  
 DB 609 NYAFVRAHSEVQSIIGQII 628

## RESULT 9

T30552  
 glucosyltransferase N - Streptococcus salivarius (fragment)  
 C:Species: Streptococcus salivarius  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30552  
 R:Jaffe, R.I.  
 submitted to the EMBL Data Library, February 1998  
 A:Description: Streptococcus salivarius VI477 gtfN.  
 A:Reference number: Z20854  
 A:Accession: T30552  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1449 <JAF>  
 A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AA05156.1  
 C:Genetics:  
 A:Gene: gtfN



Query Match 61.8%; Score 68; DB 2; Length 1449;  
Best Local Similarity 65.0%; Pred. No. 0.03;  
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARHDSVQDLIRDI 22  
:|||||:|||||:  
DB 609 NAFVRAHDSVQAVLANII 628

## RESULT 10

A44811

glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

C:Species: Streptococcus salivarius

C&gt;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999

C/Accession: A44811; S22726; S28809

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A&gt;Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A/Reference number: A44811; PMID:92148377; PMID:1838391

A/Accession: A44811

A/Molecule type: DNA

A/Residues: 1-1518 &lt;GIF&gt;

A/Cross-references: EMBL:Z11873; NID:G47526; PIDN:CAA77900.1; PID:G47527

A/Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)

C/Genetics:

A/Gene: gtfI

C/Superfamily: cpl repeat homology

C/Keywords: glycosyltransferase; hexosyltransferase

F:1307-1326/Domain: cpl repeat homology &lt;CP4&gt;

Query Match 59.1%; Score 65; DB 2; Length 1518;  
Best Local Similarity 60.0%; Pred. No. 0.094;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARHDSVQDLIRDI 22  
:|||||:|||||:  
DB 604 NYIFRAHDSVQAVLANII 623

## RESULT 11

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C&gt;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C/Accession: A45866

R:Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A&gt;Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A/Reference number: A45866; MUID:91100958; PMID:2148600

A/Accession: A45866

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1431 &lt;HON&gt;

A/Cross-references: GB:M29236

C/Superfamily: cpl repeat homology

C/Keywords: glycosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology &lt;CP1&gt;

F:1127-1146/Domain: cpl repeat homology &lt;CP2&gt;

F:1192-1211/Domain: cpl repeat homology &lt;CP3&gt;

F:1257-1276/Domain: cpl repeat homology &lt;CP4&gt;

F:1277-1297/Domain: cpl repeat homology &lt;CP5&gt;

F:1321-1340/Domain: cpl repeat homology &lt;CP6&gt;

F:1341-1361/Domain: cpl repeat homology &lt;CP6&gt;

F:1385-1404/Domain: cpl repeat homology &lt;CP7&gt;

Query Match 56.4%; Score 62; DB 2; Length 1431;  
Best Local Similarity 65.0%; Pred. No. 0.26;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARHDSVQDLIRDI 22  
:|||||:|||||:  
DB 576 NYIFRAHDSVQAVLANII 595

## RESULT 12

T30858

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C&gt;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C/Accession: T30858

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A&gt;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr

A/Reference number: Z20909; MUID:95122197; PMID:7822030

A/Accession: T30858

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1577 &lt;SIM&gt;

A/Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AMC41413.1

C/Genetics:

A/Gene: gtfm

Query Match 56.4%; Score 62; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.3;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARHDSVQDLIRDI 22  
:|||||:|||||:  
DB 661 NYIFRAHDSVQAVLANII 680

## RESULT 13

S75540

hypothetical protein s111219 - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C/Accession: S75540

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,

DNA Res. 3, 109-136, 1996

A&gt;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S75540

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-108 &lt;KAN&gt;

A/Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BA18101.1; PID:g165311

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C/Superfamily: Synecocystis hypothetical protein s111219

Query Match 44.5%; Score 49; DB 2; Length 108;  
Best Local Similarity 38.9%; Pred. No. 1.5;  
Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 5 SPARHDSVQDLIRDI 22  
:|||||:|||||:  
DB 34 NWARHDSVDVQVQETL 51

## RESULT 14

G82455

hypothetical protein VCA0471 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C/Accession: G82455

R:Heidelberg, U.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.

Chaidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers,

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Frazer, C.M.

Nature 406, 477-483, 2000

A&gt;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: G82455

A/Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-51 <HRT>
A:Cross-references: GB:AE004379; GB:AE003853; NID:g9657865; PIDN:AAF96375.1; GSPDB:GN001
C:Experimental source: BeroGroup OL; Strain NI6961; Biotype EI Tor
C:Genetics:
A:Gene: VCA0471
A:Map position: 2

Query Match          42.7% Score 47; DB 2; Length 51;
Best Local Similarity 50.0%; Pred. No. 1.3;
Matches 11; Conservative 6; Mismatches 3; Indels 2; Gaps 1;

QY      1 VPYSFPAHDSFYODLRDII 22
        |||::|||::|||::|||
Db       15 VPAYSAIR-NSEIRDLERKII 34

RESULT 15
R83786
galactose-1-phosphate uridylyltransferase galT [imported] - Bacillus halodurans (strain C
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 27-Oct-2003
C:Accession: R83786
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058332
A:Accession: R83788
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:gl0173440; PIDN:BAB04828.1; GSPDB:GN000
C:Experimental source: strain C-125
C:Genetics:
A:Gene: galT
C:Superfamily: galactose-1-phosphate uridylyltransferase, Bacillus type

Query Match          42.7% Score 47; DB 2; Length 508;
Best Local Similarity 56.2%; Pred. No. 19;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      7 AAHADSEYODLRDII 22
        |||::|||::|||::|||
Db       460 AHLDSDNYMDLDRDI 475

RESULT 16
P82935
DNA polymerase III gamma-tau subunits U0087 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Update: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: P82935
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mini
A:Reference number: AB2870
A:Accession: P82935
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-597 <GUA>
A:Cross-references: GB:AF002108; GB:AF222894; NID:g6899034; PIDN:AAF30492.1; GSPDB:GN001
C:Experimental source: BeroVar 3; Biovar 1
C:Genetics:
A:Gene: dnaX; U0087
A:Genetic code: SGC3

Query Match          42.7% Score 47; DB 2; Length 597;
Best Local Similarity 47.4%; Pred. No. 23;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      4 YSFARAHDSFYODLRDII 22
        |||::|||::|||::|||
Db       173 YDFKRLNSELOELIDSLI 191

```

RESULT 17  
 AB1229  
 excinuclease ABC chain C homolog uvrc [imported] - *Listeria monocytogenes* (strain EGD-e)  
 C:Species: *Listeria monocytogenes*  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C:Accession: AB1229  
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurgat, O.; Entlian, K.D.; Fsihl, H.  
 Science 294, 849-852, 2001  
 A:Authors: Kretz, U.; Kuhn, M.; Kunst, F.; Kurupkat, G.; Madeno, E.; Maitournam, A.; Ma  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A.:Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; PMID:21537279; PMID:11679669  
 A:Accession: AB1229  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-603 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAC99312.1; PID: g16410650; GSPDB:GN00177  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: uvrc  
 C:Superfamily: excinuclease ABC chain C  
  
 Query Match 42.3%; Score 46.5; DB 2; Length 603;  
 Best Local Similarity 45.5%; Pred. No. 28;  
 Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;  
  
 QY 1 VPSYSPARAHSEVODLIRDI 22  
 DB 42 VRSY-PSGTHDSKTQRLVQETV 62  
  
 RESULT 18  
 AD1582  
 excinuclease ABC chain C homolog uvrc [imported] - *Listeria innocua* (strain Clp11262)  
 C:Species: *Listeria innocua*  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C:Accession: AD1582  
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurgat, O.; Entlian, K.D.; Fsihl, H.  
 Science 294, 849-852, 2001  
 A:Authors: Kretz, U.; Kuhn, M.; Kunst, F.; Kurupkat, G.; Madeno, E.; Maitournam, A.; Ma  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A.:Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; PMID:21537279; PMID:11679669  
 A:Accession: AD1582  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-603 <GLA>  
 A:Cross-references: GB:AL592022; PIDN:CAC96428.1; PID: g16413656; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Gene: uvrc  
 C:Superfamily: excinuclease ABC chain C  
  
 Query Match 42.3%; Score 46.5; DB 2; Length 603;  
 Best Local Similarity 45.5%; Pred. No. 28;  
 Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;  
  
 QY 1 VPSYSPARAHSEVODLIRDI 22  
 DB 42 VRSY-PSGTHDSKTQRLVQETV 62  
  
 RESULT 19  
 GNPF17  
 retrovirus-related pol polypeptide - fruit fly (*Drosophila melanogaster*) retrotransposon  
 A:Alternate names: reverse transcriptase  
 C:Species: *Drosophila melanogaster*

C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #extl\_change 16-Jun-2000  
C/Accession: A033971  
R/Saigo, K.; Kugimiyu, W.; Matsuo, Y.; Inouye, S.; Yoshioka, K.; Yuki, S.  
Nature 312, 659-661, 1984  
A/Title: Identification of the coding sequence for a reverse transcriptase-like enzyme  
A/Reference number: A93349; PMID:85061628; PMID:6209583  
A/Accession: A033971  
A/Molecule type: DNA  
A/Residues: 1-1058 <SAT>  
A/Cross-references: GB:X01472; GB:J01060; GB:J01061; NID:98142; PIDD:CAA25702.1; PID:913  
C/Genetic: 1  
A/Genes: PLYBase:17.6  
A/Cross-references: PLYBase:FBgm0000004  
C/Superfamily: pol polyprotein  
C/Keywords: polyprotein; reverse transcriptase

Query Match 40.9%; Score 45; DB 1; Length 1058;  
Best Local Similarity 36.8%; Pred. No. 93;  
Matches 7; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 4 YSFARAHSEVQDLIRDI 22  
Db 213 YSYQAYEQEVESQIDML 231

RESULT 20  
F69471  
atrazine chlorohydrolyase homolog - Archaeoglobus fulgidus  
C/Species: Archaeoglobus fulgidus  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #extl\_change 22-Oct-1999  
C/Accession: F69471  
R/Klen, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A/Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaime, B.F.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.; Saito, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A/Reference number: A69250; PMID:98049343; PMID:9389475  
A/Accession: F69471  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-330 <KLE>  
A/Cross-references: GB:AE000980; GB:AE000782; NID:92689303; PIDD:AA889475.1; PID:9264877

Query Match 40.0%; Score 44; DB 2; Length 330;  
Best Local Similarity 42.1%; Pred. No. 35;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDLIRDI 21  
Db 152 AYSSARDHDKMEVEREI 170

RESULT 21  
T47184  
hypothetical protein DKFZp434F1526.1 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #extl\_change 20-Apr-2000  
C/Accession: T47184  
R/Amstrong, W.; Wikner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A/Reference number: Z24375  
A/Accession: T47184  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-506 <AAA>  
A/Cross-references: EMBL:AL162051  
A/Experimental source: adult testis; clone DKFZp434F1526  
C/Genetics: 1  
A/Note: DKFZp434F1526.1

Query Match 40.0%; Score 44; DB 2; Length 506;

Best Local Similarity 31.8%; Pred. No. 57;  
Matches 7; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPSIFARAHSEVQDLIRDI 22  
Db 456 VQYFYSNLPSPESKDIIQQOV 477

RESULT 22  
OHUCR  
natriuretic peptide receptor C precursor - human  
N/Alternative names: atrial natriuretic factor clearance receptor; natriuretic peptide cl  
C/Species: Homo sapiens (man)  
C/Date: 09-Nov-1990 #sequence\_revision 05-May-1995 #extl\_change 22-Jun-1999  
C/Accession: S10150; A35896  
R/Lowe, D.G.; Camerato, T.R.; Goeddel, D.V.  
Nucleic Acids Res. 18, 3412, 1990  
A/Title: cDNA sequence of the human atrial natriuretic peptide clearance receptor.  
A/Reference number: S10150; PMID:90287735; PMID:2162522  
A/Accession: S10150  
A/Molecule type: mRNA  
A/Residues: 1-540 <LOW>  
A/Cross-references: EMBL:X52282; NID:928705; PIDD:CAA36523.1; PID:928706  
A/Note: alternative splice form C6  
R/Porter, J.G.; Arfsten, A.; Fuller, F.; Miller, J.A.; Gregory, L.C.; Lewicki, J.A.  
Biochem. Biophys. Res. Commun. 171, 796-803, 1990  
A/Title: Isolation and functional expression of the human atrial natriuretic peptide cl  
A/Reference number: A35896; PMID:90386656; PMID:2169733  
A/Accession: A35896  
A/Molecule type: mRNA  
A/Residues: 1-475 'SG', 477-540 <POR>  
A/Cross-references: GB:M59305; NID:9178651; PIDD:AA51734.1; PID:9178652  
A/Note: alternative splice form C5  
R/Stults, J.T.; O'Connell, K.L.; Garcia, C.; Wong, S.; Engel, A.M.; Garbers, D.L.; Lowe,  
Biochemistry 33, 11372-11381, 1994  
A/Title: The disulfide linkages and glycosylation sites of the human natriuretic peptid  
A/Reference number: A58870; PMID:95244450; PMID:7727388  
A/Contents: annotation; propeptide, disulfide bonds, carbohydrate binding sites  
C/Genetics: 1  
A/Genes: GDB:NR3, NRRC, ANRC  
A/Cross-references: GDB:125201; OMIM:108962  
A/Map position: 5p14-5p13  
C/Superfamily: natriuretic peptide receptor C; natriuretic peptide-binding domain homol  
C/Keywords: alternative splicing; glycoprotein; homodimer; hormone receptor; transmembr  
F/1-23/Domain: signal sequence #status predicted <SIG>  
F/24-45/Domain: propeptide #status experimental <PRO>  
F/46-540/Product: natriuretic peptide receptor C #status predicted <MAT>  
F/46-479/Domain: extracellular #status predicted <EXT>  
F/131-462/Domain: natriuretic peptide-binding domain homology <NPB>  
F/483-503/Domain: transmembrane #status predicted <TM>  
F/86-293/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F/108-136, 213-261/Disulfide bonds: #status experimental  
F/394/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental  
F/473-476/Disulfide bonds: (partial) (in splice form C6) #status experimental

Query Match 40.0%; Score 44; DB 1; Length 540;  
Best Local Similarity 38.9%; Pred. No. 62;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 4 YSFARAHSEVQDLIRDI 21  
Db 234 YSFDETDLDLIRVNI 251

RESULT 23  
T16450  
hypothetical protein F53B3.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #extl\_change 20-Sep-1999  
C/Accession: T16450  
R/Miller, N.  
submitted to the EMBL Data Library, November 1995  
A/Description: The sequence of C. elegans cosmid F53B3.

A:Reference number: Z18515  
A:Accession: F16450  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-651 <ML>  
A:Cross-references: EMBL:U00414; NID:g1065474; PID:g1065476; PIDN:AAA81405.1; CESP:F53B3  
C:Genetics:  
A:Gene: CESP:F53B3.2  
A:Initrns: 76/1; 206/1; 307/2; 384/3; 412/1; 450/3; 492/1; 600/3; 633/3  
Query Match 40.0%; Score 44; DB 2; Length 651;  
Best Local Similarity 47.1%; Pred. No. 77;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 4 YSFARHDSVODLIRD 20  
DB 435 YSHSTWIDNSIQDLIRD 451  
RESULT 24  
T31679  
Bacitracin synthetase 3 - Bacillus licheniformis  
C:Species: Bacillus licheniformis  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T31679  
R:Konz, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M.A.  
Chem. Biol. 4, 927-937, 1997  
A:Title: The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecu  
A:Reference number: Z21058; MUID:98089193; PMID:9427658  
A:Accession: T31679  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-6359 <KON>  
A:Cross-references: EMBL:AF007865; NID:g4464275; PID:g2982196; PIDN:AAC06348.1  
C:Genetics:  
A:Gene: bacC  
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology  
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F:505-946/Domain: acetate-CoA ligase homology <ACLI>  
F:964-1032/Domain: acyl carrier protein homology <ACPI>  
F:1542-1978/Domain: acetate-CoA ligase homology <ACI2>  
F:1996-2063/Domain: acyl carrier protein homology <ACPI2>  
F:3046-3483/Domain: acetate-CoA ligase homology <ACI3>  
F:3500-3568/Domain: acyl carrier protein homology <ACPI3>  
F:4078-4526/Domain: acetate-CoA ligase homology <ACI4>  
F:4542-4609/Domain: acyl carrier protein homology <ACPI4>  
F:5593-6032/Domain: acetate-CoA ligase homology <ACI5>  
F:6050-6118/Domain: acyl carrier protein homology <ACPI5>  
F:3532/Binding site: phosphopantetheine (Ser) (covalent) #status predicted  
Query Match 40.0%; Score 44; DB 2; Length 6359;  
Best Local Similarity 45.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 4; Mismatches 4; Indels 4; Gaps 1;  
QY 1 VPSYFARHDSVODLIRDII 22  
DB 3186 VPSFSF---DSSVSDIFTTII 3203  
RESULT 25  
A89820  
glutamyl-tRNA synthetase [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: A89820  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguno, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C:Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: A89820  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-484 <KUR>  
A:Cross-references: GB:BA000018; PID:g13700418; PIDN:BA81716.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: gltX  
C:Superfamily: glutamate-tRNA ligase; glutamine-tRNA ligase homology  
Query Match 39.5%; Score 43.5; DB 2; Length 484;  
Best Local Similarity 40.9%; Pred. No. 65;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;  
QY 1 VPSYFARA---HDSVODLIR 19  
DB 193 IPTYFRAVALDDHYMQISDVIR 214  
RESULT 26  
F83819  
hypothetical protein BHJ358 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83819  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: F83819  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-176 <STO>  
A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA805077.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BHJ358  
Query Match 39.1%; Score 43; DB 2; Length 176;  
Best Local Similarity 38.5%; Pred. No. 24;  
Matches 10; Conservative 3; Mismatches 7; Indels 6; Gaps 1;  
QY 1 VPSYFARHDS-----VODLIRD 20  
DB 83 VESYDYEDLDHRCQOTQPTVQDSVRD 108  
RESULT 27  
F72718  
hypothetical protein APE0287 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: F72718  
R:Kawabazaki, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: F72718  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <KAN>  
A:Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BA879242.1; PID:g5103926  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0287  
C:Superfamily: Aeropyrum pernix hypothetical protein APE0287  
Query Match 39.1%; Score 43; DB 2; Length 195;  
Best Local Similarity 31.8%; Pred. No. 27;  
Matches 7; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
QY 1 VPSYFARHDSVODLIRDII 22  
DB 6 IPGYDYGVERKSPITDLRFLL 27

## RESULT 28

T50180

nucleolar protein NOP5-like protein [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50180  
R:Radcock, K.; Church, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
Submitted to the EMBL Data Library, February 2000  
A:Reference number: Z25044  
A:Accession: T50180  
A:Status: preliminary; translated from GB/EMBL/DBSJ  
A:Molecule type: DNA  
A:Residues: 1-508 <BAD>  
A:Cross-references: EMBL:AL138854; PIDN:CA872231.1; GSPDB:GN00066; SPDB:SPAC23G3.06  
C:Experimental source: strain 972h(-); cosmid c23G3  
C:Genetics:  
A:Gene: SPDB:SPAC23G3.06  
A:Map position: 1  
A:Introns: 6/1; 40/2  
C:Superfamily: garden pea SAR DNA-binding protein

Query Match 39.1%; Score 43; DB 2; Length 508;  
Best Local Similarity 47.6%; Pred. No. 83;  
Matches 10; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDLIRDI 21

Db 100 LFGLEFRIISDSVQDLVRGI 120

## RESULT 29

F70332

proline-tRNA synthetase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999

C:Accession: F70332

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:9819666; PMID:9537320  
A:Accession: F70332  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-570 <AQF>  
A:Cross-references: GB:AE00066; NID:G2983038; PIDN:AA06648.1; PID:G2983039; GB:AE00065  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: proS  
C:Superfamily: proline-tRNA ligase

Query Match 39.1%; Score 43; DB 2; Length 570;  
Best Local Similarity 31.6%; Pred. No. 95;  
Matches 6; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 4 YSPARAHSEVQDLIRDI 22

Db 104 YCLGPTHERITDLVRRI 122

## RESULT 30

H90599

conserved hypothetical protein MYPV\_7040 [imported] - Mycoplasma pulmonis (strain UAB CT  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: H90599  
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: H90599

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:02:41 ; Search time 8 Seconds

(Without alignments)  
145.193 Million cell updates/sec

Title: US-09-290-049A-17

Perfect score: 110  
Sequence: 1 VPSISFARAHSEVQDLIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1597	1	GTF1_STRDO
2	108	98.2	1592	1	GTF2_STRDO
3	105	95.5	1476	1	GTFB_STRMT
4	100	90.9	1455	1	GTRC_STRMT
5	70	63.6	1365	1	GTRF_STRMT
6	62	56.4	1462	1	GTFD_STRMT
7	47	42.7	508	1	GALT_BACHD
8	46.5	42.3	603	1	UVRC_LISIN
9	46.5	42.3	603	1	UVRC_LISMO
10	45	40.9	258	1	UBIR_RHIC
11	45	40.9	993	1	GCSF_XILFT
12	45	40.9	1058	1	POJ3_DROME
13	44.5	40.5	484	1	SYE_STRAP
14	44	40.0	330	1	YH75_ARCFU
15	44	40.0	541	1	ANPC_HUMAN
16	44	40.0	6359	1	BACC_BACCI
17	43.5	39.5	484	1	SYE_STRAM
18	42.5	38.6	486	1	SYE_STRAP
19	42.5	38.6	540	1	HXTD_YEAST
20	42	38.2	196	1	MAD2_YEAST
21	42	38.2	280	1	YIBO_HAEIN
22	42	38.2	429	1	TF3A_YEAST
23	42	38.2	1046	1	RPOC_WEIHE
24	42	38.2	1363	1	PRAX_RAT
25	42	38.2	1835	1	DURI_YEAST
26	42	38.2	295	1	FMRE_LYMST
27	41	37.3	306	1	MITF_MOUSE
28	41	37.3	526	1	EHD4_HUMAN
29	41	37.3	541	1	EHD4_MOUSE
30	41	37.3	541	1	FUMA_ECOLI
31	41	37.3	541	1	WR20_ARATH
32	41	37.3	579	1	FUMA_SALTY
33	41	37.3	579	1	FUMA_SALTY

34	41	37.3	799	1	SYL_MYCPE	Q8W18 mycoplasma
35	41	37.3	933	1	ODOI_ECOLI	P07015 escherichia
36	41	37.3	953	1	COPE_HUMAN	P53618 mus sapien
37	41	37.3	953	1	COPE_MOUSE	Q911f7 mus musculu
38	41	37.3	953	1	COPE_RAT	P23514 rattus norv
39	40.5	36.8	196	1	YF71_CAEEL	Q09220 caenorhabdi
40	40.5	36.8	395	1	VORA_PYRAB	Q94921 pyrococcus
41	40.5	36.8	484	1	SYE_MYCPN	P75114 mycoplasma
42	40.5	36.8	124	1	RBS_PSEHY	Q51857 pseudomonas
43	40	36.4	179	1	DEP2_AMASP	Q84971 anabaena sp
44	40	36.4	344	1	TOB1_SEVRA	P16472 shope fibro
45	40	36.4	344	1	MI2D_BACSU	P26935 bacillus su
46	40	36.4	387	1	DEP1_MOUSE	Q94966 mus musculu
47	40	36.4	397	1	DEP1_RAT	P56163 rattus norv
48	40	36.4	419	1	ARG2_RHIME	Q92965 rhizobium m
49	40	36.4	495	1	Y192_MYCTU	Q10532 mycobacteri
50	40	36.4	537	1	ANPC_BOVIN	P10730 bos taurus
51	40	36.4	603	1	PRIM_BACSU	P05096 bacillus su
52	40	36.4	609	1	YL15_MYCLE	P46509 mycobacteri
53	40	36.4	638	1	HTPG_STRCO	P58481 streptomyce
54	40	36.4	919	1	Y193_HUMAN	Q94967 homo sapien
55	40	36.4	993	1	GCSF_XILFT	Q9pdj4 xyliella fas
56	40	36.4	1068	1	P11A_BOVIN	P32871 bos taurus
57	40	36.4	1068	1	P11A_HUMAN	P42336 homo sapien
58	40	36.4	1068	1	P11A_MOUSE	P42337 mus musculu
59	40	36.4	1418	1	SMC4_YEAST	Q12267 saccharomyc
60	40	36.4	1452	1	VG12_YEAST	P10033 feline infe
61	40	36.4	1461	1	PRAX_HUMAN	Q9bmo homo sapien
62	40	36.4	1690	1	C190_DROME	Q94761 dirosophila
63	40	35.9	489	1	SYE_OCEBI	Q8e02 oceanobacil
64	39.5	35.9	759	1	SC71_YEAST	P32784 saccharomyc
65	39	35.5	190	1	CYC6_CYACA	Q9t1w1 cyanidium c
66	39	35.5	190	1	DDC_THLTV	Q823k6 chlamydomon
67	39	35.5	273	1	PAT5_METVA	Q58378 methanococc
68	39	35.5	282	1	Y1AJ_ECOLI	P37671 escherichia
69	39	35.5	309	1	EFPS_RICPR	Q92660 neisseria g
70	39	35.5	310	1	PIP_NEIGO	P42786 neisseria g
71	39	35.5	414	1	FTZ2_PYRHO	Q58491 pyrococcus
72	39	35.5	453	1	MDHP_FLABI	P46489 flaveria bl
73	39	35.5	540	1	SCX_HUMAN	P98077 homo sapien
74	39	35.5	588	1	SYD_XANAC	Q8pms xanthomonas
75	39	35.5	588	1	SYD_XANCP	Q8p60 xanthomonas
76	39	35.5	604	1	VER1_HPV36	P50808 human papil
77	39	35.5	609	1	YL15_MYCTU	Q33250 mycobacteri
78	39	35.5	624	1	SIR_SYNP7	P30008 mycobacteri
79	39	35.5	656	1	UVAB_MYCCE	P47319 mycoplasma
80	39	35.5	691	1	Y104_YEAST	P40460 saccharomyc
81	39	35.5	932	1	H1RL_SCHPO	P87314 schizosaccha
82	39	35.5	1018	1	M2B2_MOUSE	Q54782 mus musculu
83	39	35.5	1249	1	APAF_RAT	Q9epv5 rattus norv
84	39	35.5	1365	1	KRES_YEAST	P22023 saccharomyc
85	39	35.5	1391	1	PRAX_MOUSE	Q55103 mus musculu
86	39	35.5	1447	1	VG12_CVPRU	P07946 porcine tra
87	39	35.5	1447	1	VG12_CVPRU	Q01977 porcine tra
88	39	35.5	1451	1	VG12_CVCAI	P36300 canine ente
89	39	35.5	2541	1	T1N1_HUMAN	Q94900 homo sapien
90	38.5	35.0	348	1	LPXK_AGRTS	Q8u15 agrobacteri
91	38.5	35.0	364	1	SERC_DROME	Q9van0 drosophila
92	38.5	35.0	468	1	ST11_SCHPO	P36631 schizosaccha
93	38.5	35.0	481	1	SYE_STRP3	Q8Xp93 streptococc
94	38.5	35.0	481	1	SYE_STRP3	Q9a1j8 streptococc
95	38.5	35.0	483	1	SYE_LACIA	Q9c4e7 lactococcus
96	38.5	35.0	484	1	SYE_STRAS	Q8e4p2 streptococc
97	38.5	35.0	484	1	SYE_STRAS	Q8e4x4 streptococc
98	38.5	35.0	486	1	SYE_STRMN	Q8d88 streptococc
99	38.5	35.0	486	1	SYE_STRMN	Q979g1 streptococc
100	38.5	35.0	598	1	UVRC_BACSU	P14951 bacillus su

## ALIGNMENTS

RESULT 1

STFD	STROD	STANDARD	PRT	1597 AA.
GTFL	STROD			
AC	PI1001;			
AD	01-JUL-1989 (Rel. 11, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)			
DE	(Sucrose 6-glucosyltransferase).			
GN	GTF1.			
OS	Streptococcus downei (Streptococcus sobrinus).			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
CC	Streptococcus.			
CC	NCBI_TaxID=1317;			
CC	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=MFE28;			
RX	MEDLINE=87308014; PubMed=3040686;			
RA	Ferretti J.J., Gilpin M.L., Russell R.B.;			
RT	"Nucleotide sequence of a glucosyltransferase gene from Streptococcus			
RT	sobrinus Mfe28."			
RL	J. Bacteriol. 169:4271-4278 (1987).			
CC	-1 FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT			
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE			
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE			
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	-1 CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl) (N) = D-			
CC	fructose + ((1,6)-alpha-D-glucosyl) (N+1).			
CC	-1 SUBCELLULAR LOCATION: Secreted.			
CC	-1 MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA			
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES			
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOS). GTF-S1 SYNTHESIZES BOTH			
CC	FORMS OF GLUCANS.			
CC	-1 SIMILARITY: Belongs to family 70 of glycosyl hydrolases.			
CC	-1 SIMILARITY: Contains 19 cell wall binding repeats.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb.ch">license@isb.ch</a> ).			
CC	-----			
DR	EMBL; M17391; AAC63063.1; -			
DR	InterPro; IPR002479; CM binding.			
DR	InterPro; IPR003318; Glyco binding_70.			
DR	Pfam; PF01473; CM_binding_1; 16.			
KW	transferase; Glycosyltransferase; Signal; Repeat; Dental caries.			
FT	SIGNAL	1	1597	POTENTIAL.
FT	CHAIN	39	1050	GLUCOSYLTRANSFERASE-1.
FT	DOMAIN	1039	1597	CATALYTIC (APPROXIMATE).
FT	DOMAIN	1039	1597	GLUCAN-BINDING (APPROXIMATE).
FT	REPEAT	1039	1132	1.25 A, 2 B, AND 5 AC REPEATS.
FT	REPEAT	1163	1213	A REPEAT.
FT	REPEAT	1227	1277	AC REPEAT.
FT	REPEAT	1292	1342	AC REPEAT.
FT	REPEAT	1352	1399	B REPEAT.
FT	REPEAT	1406	1455	AC REPEAT.
FT	REPEAT	1465	1512	B REPEAT.
FT	REPEAT	1519	1568	AC REPEAT.
FT	REPEAT	1582	1597	A REPEAT (INCOMPLETE).
SEQ	SEQUENCE	1597 AA;	177080 MM;	B9E6A200686799E CRC64;

Query Match	100.0%;	Score 110;	Length 1587;
Best Local Similarity	100.0%;	Pred. No. 2.1e-09;	
Matches	22;	Conservative	0;
		Mismatches	0;
		Indels	0;
		Gaps	0;

```
QY      1 VPSYSFARAHDSFVQDLIRDTI 22
        |||||
Db      554 VPSYSFARAHDSFVQDLIRDTI 575
```

	RESULT 2
CC	GTF2_STRDO
ID	_GTF2_STRDO
STANDARD;	PRT; 1592 AA.
AC	P27470;
DT	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DS	(Stucose 6-glucosyltransferase).
OS	Streptococcus downei (Streptococcus sobrinus).
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=13117;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=6715 / Serotype G;
RX	MEDLINE=91123227; PubMed=1704006;
RA	Abo H., Matsumura T., Kodama T., Onha H., Fukui K., Kato K.,
RA	Kagawa H.;
RT	"Peptide sequences for sucrose splitting and glucan binding within
RT	Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT	synthetase)",
RL	J. Bacteriol. 173:989-996(1991).
CC	-I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC	-I- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
CC	fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC	-I- SUBCELLULAR LOCATION: Secreted
CC	-I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S1 SYNTHESIZES BOTH
CC	FORMS OF GLUCANS.
CC	-I- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.
CC	-I- SIMILARITY: Contains 16 cell wall binding repeats.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; D90213; BAAL4241.1; -
DR	InterPro; IPR002479; CW binding.
DR	InterPro; IPR003318; Glyco_hydro_70.
DR	Pfam; PF01473; CW binding_I; 13.
DR	Pfam; PF02324; Glyco_hydro_70_1.
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental carrier.
FT	SIGNAL 1 38 POTENTIAL.
FT	CHAIN 39 1592
FT	DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
FT	DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
FT	REPEAT 1093 1592 6.5 X TANDEM REPEATS.
FT	REPEAT 1093 1142 1.
FT	REPEAT 1158 1207 2.
FT	REPEAT 1222 1272 3.
FT	REPEAT 1287 1337 4.
FT	REPEAT 1402 1451 5.
FT	REPEAT 1514 1563 6.
FT	REPEAT 1577 1592 7 (INCOMPLETE).
SEQ	SEQUENCE 1592 AA; 176167 MW; BC0A65D079351BCF CR664;

Query Match	98.28;	Score 108;	DB 1;	Length 152;
Best Local Similarity	95.58;	Pred. No. 4.5e-09;		
Matches 21; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      1 VPSYSFARAHDSVQDLIRDI 22
        |||||:|||||
Db      548 VPSYSFARAHDSVQDLIRDI 565
```







DB 578 VPSYSPRAHDSVQDLIRNII 599

## RESULT 5

GTS\_STRDO STANDARD; PRT; 1365 AA.

AC P29336;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
 DE (Sucrose 6-glucosyltransferase).  
 GN GTF-S.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MFE28;  
 RX MEDLINE=90316665; PubMed=2142479;  
 RA Gilmore K.S., Russell R.R., Ferretti J.J.;  
 RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
 RT glucosyltransferase that synthesizes soluble glucans."  
 RT Infect. Immun. 58:2452-2458 (1990).  
 RL [1]  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-  
 CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).  
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
 CC PRIMER GLUCAN UNLIKE GTF-1.  
 CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
 CC 1,6-GLUCOSE).  
 CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
 CC -1- SIMILARITY: Contains 10 cell wall binding repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; M30943; AAA26898.1; -  
 CC InterPro; IPR002479; CW binding.  
 CC InterPro; IPR003318; Glyco\_hydro\_70.  
 CC Pfam; PF01473; CW binding\_1; 8.  
 CC Pfam; PF00324; Glyco\_hydro\_70; 1.  
 CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 CC SIGNAL 1 36  
 CC OR 37 (POTENTIAL).  
 CC FT CHAIN 37 1365 GLUCOSYLTANSFERASE-S.  
 CC FT REPEAT 157 177 CELL WALL BINDING 1.  
 CC FT REPEAT 178 197 CELL WALL BINDING 2.  
 CC FT DOMAIN 198 1061 CATALYTIC (APPROXIMATE).  
 CC FT REPEAT 1062 1082 CELL WALL BINDING 3.  
 CC FT REPEAT 1083 1102 CELL WALL BINDING 4.  
 CC FT REPEAT 1150 1169 CELL WALL BINDING 5.  
 CC FT REPEAT 1170 1190 CELL WALL BINDING 6.  
 CC FT REPEAT 1225 1243 CELL WALL BINDING 7.  
 CC FT REPEAT 1289 1308 CELL WALL BINDING 8.  
 CC FT REPEAT 1309 1328 CELL WALL BINDING 9.  
 CC FT REPEAT 1331 1352 CELL WALL BINDING 10.  
 CC SEQUENCE 1365 AA; 151590 MW; 16729685A2E8C476 CRC64;

Query Match 63.6%; Score 70; DB 1; Length 1365;

Best Local Similarity 68.2%; Pred. 0.0048;

Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

1 VPSYSPRAHDSVQDLIRNII 22

DB 537 VENVYPIRAHDSRVQTRIAKII 558

## RESULT 6

GTFD\_STRMU STANDARD; PRT; 1462 AA.

AC P49331; 069383; 069386; 069389; 069392; 069398;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
 DE (Sucrose 6-glucosyltransferase).  
 GN GTFD OR SMU.910.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RX MEDLINE=91100958; PubMed=2148600;  
 RA Honda O., Kato C., Kuramitsu H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 RT the glucosyltransferase-S enzyme."  
 RT J. Gen. Microbiol. 136:2099-2105 (1990).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,  
 RC MT4467 / Serotype B, and MT8148 / Serotype C;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujisawa T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans."  
 RT FEMS Microbiol. Lett. 161:331-336 (1998).  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=UA159 / ATCC 700610 / Serotype C;  
 RC MEDLINE=22295063; PubMed=12397186;  
 RA Adic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 RT pathogen."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).  
 RL [1]  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-  
 CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
 CC -1- SIMILARITY: Contains 6 cell wall binding repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; M29296; AAA26895.1; -  
 CC EMBL; D88653; BAA26103.1; -  
 CC EMBL; D88656; BAA26107.1; -  
 CC EMBL; D88659; BAA26111.1; -  
 CC EMBL; D88662; BAA26115.1; -  
 CC EMBL; D89979; BAA26121.1; -

DR EMBL: AE014932; AAN58619.1; -.  
 DR InterPro: IPR002479; CM binding\_70.  
 DR pfam: PF01473; CM\_binding\_11.  
 DR pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase; Glycosyltransferase; Signal; Repeat; Dental caries;  
 KW Complete proteome.  
 FT SIGNAL 1 1462 ?  
 FT CHAIN 1 1462 ?  
 FT DOMAIN 1232 1433  
 FT REPEAT 1232 1295  
 FT REPEAT 1296 1359  
 FT REPEAT 1360 1423  
 FT REPEAT 10 10  
 FT VARIANT 19 19  
 FT VARIANT 58 58  
 FT VARIANT 68 68  
 FT VARIANT 81 81  
 FT VARIANT 113 113  
 FT VARIANT 122 122  
 FT VARIANT 132 132  
 FT VARIANT 135 135  
 FT VARIANT 137 137  
 FT VARIANT 202 202  
 FT VARIANT 255 255  
 FT VARIANT 275 275  
 FT VARIANT 288 288  
 FT VARIANT 301 301  
 FT VARIANT 313 313  
 FT VARIANT 317 317  
 FT VARIANT 328 328  
 FT VARIANT 350 350  
 FT VARIANT 628 628  
 FT VARIANT 688 688  
 FT VARIANT 726 726  
 FT VARIANT 762 762  
 FT VARIANT 964 964  
 FT VARIANT 1019 1019  
 FT VARIANT 1059 1059  
 FT VARIANT 1060 1060  
 FT VARIANT 1080 1080  
 FT VARIANT 1142 1142  
 FT VARIANT 1198 1198  
 FT VARIANT 1220 1220  
 FT VARIANT 1280 1280  
 FT VARIANT 1282 1282  
 FT VARIANT 1290 1290  
 FT VARIANT 1311 1311  
 FT VARIANT 1403 1403  
 FT VARIANT 1425 1425  
 FT VARIANT 1449 1449  
 FT CONFLICT 1428 1462  
 SQ SEQUENCE 1462 AA; 163387 MW; CEAA279CAD708645 CRC64;  
 Query Match 56.4%; Score 62; DB 1; Length 1462;  
 Best Local Similarity 65.0%; Pred. No. 0.1;  
 Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

RESULT 7  
 GALT\_BACHD STANDARD; PRT; 508 AA.  
 ID GALT\_BACHD: Q9RC74;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) (Gal-1-P  
 uridylyltransferase) (UDP-glucose--hexose-1-phosphate  
 uridylyltransferase).  
 GN GALT OR BH109.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NCBI\_TaxID:86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 RP [2]  
 RP SEQUENCE OF 1-450 FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=99411980; PubMed=10484179;  
 RA Takami H., Takaki Y., Nakasone K., Sakiyama T., Maeno G., Sasaki R.,  
 RA Hirama C., Fuji F., Masui N.;  
 RT "Genetic analysis of the chromosome of alkaliphilic Bacillus  
 RT halodurans C-125";  
 RL Extremophiles 3:227-233(1999).  
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + alpha-D-galactose 1-phosphate =  
 CC alpha-D-glucose 1-phosphate + UDP-galactose.  
 CC -1- PATHWAY: galactose metabolism, second step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: Belongs to the galactose-1-phosphate  
 CC uridylyltransferase family 2.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AP001510; BAB04828.1; -.  
 CC EMBL: AB024554; BAB83925.1; -.  
 CC PIR: E83788; E83788.  
 CC HAMAP: MF\_00571; -. 1.  
 DR InterPro: IPR000766; GalP transf II.  
 DR InterPro: IPR005850; GalP transf C.  
 DR InterPro: IPR005849; GalP transf N.  
 DR InterPro: IPR005934; GalT2.  
 DR Pfam: PF02744; GalP\_UDP\_tr\_C\_1.  
 DR Pfam: PF01087; GalP\_UDP\_transf\_1.  
 DR TIGRFAMs: TIGR01239; galt\_2; 1.  
 DR PROSITE: PS01163; GAL\_P\_UDP\_TRANSF II; 1.  
 KW Transferase; Nucleotidyltransferase; Galactose metabolism;  
 KW Complete proteome.  
 SQ SEQUENCE 508 AA; 57989 MW; 16A9F607FCEAE8 CRC64;  
 Query Match 42.7%; Score 47; DB 1; Length 508;  
 Best Local Similarity 56.2%; Pred. No. 7.9;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

RESULT 8
UVR_C LISIN STANDARD; PRT; 603 AA.
AC Q92CH5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE UVRABC system protein C (UVR_C protein) (Excinuclease ABC subunit C).
GN UVR_C OR L1N1197.
OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A. / Serovar 6a;
RC STRAIN=CLIP 11262; PubMed=11679669;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Ruenick C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Checouani F., Couve E., de Darvar A., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomics of Listeria species.";
RT Science 294:849-852(2001).
RL
CC -1- FUNCTION: The UVRABC repair system catalyzes the recognition and
CC processing of DNA lesions. UVR_C both incises the 5' and 3' sides
CC of the lesion. The N-terminal half is responsible for the 3'
CC incision and the C-terminal half is responsible for the 5'
CC incision (By similarity).
CC -1- SUBUNIT: Interacts with UVR_B in an incision complex (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the uvr_C family.
CC -1- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL59167; CAC96428.1; -.
CC PIR; A1582; A1582.
CC Listlist; L1N01197; -.
CC HAMAP; MF_00203; -.
CC InterPro; IPR003583; HHH 1.
CC InterPro; IPR001943; UVRB/C.
CC InterPro; IPR004791; UVR_C.
CC InterPro; IPR001162; UVR_C.
CC InterPro; IPR000305; UVR_C.
CC Pfam; PF01541; Excl_endo_N; 1.
CC Pfam; PF02151; UVR; 1.
CC ProDom; PD005870; UVR_C; 1.
CC SMART; SM00465; G1YC; 1.
CC SMART; SM00278; HhH1; 1.
CC TIGRfam; TIGR00194; uvr_C; 1.
CC PROSITE; PS50151; UVR; 1.
CC PROSITE; PS50164; UVR_C_1; 1.
CC PROSITE; PS50165; UVR_C_2; 1.
CC SOS response; Excision nuclease; DNA repair; DNA recombination;
CC DNA excision; Complete proteome.
CC DOMAIN 197 232 UVR.
CC SEQUENCE 603 AA; 69302 MW; 8FA17669DB8EAF1 CRC64;
Query Match 42.3%; Score 46.5; DB 1; Length 603;

```

```

Best Local Similarity 45.5%; Pred. No. 12;
Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Cy 1 PVSYPARADSEWDLIRDI 22
Db 42 VRSY-PSGTHDSKTRLRVQRI 62
RESULT 9
UVR_C LISMO STANDARD; PRT; 603 AA.
AC O6Y7BQ;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE UVRABC system protein C (UVR_C protein) (Excinuclease ABC subunit C).
GN UVR_C OR LMO1234.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A. / Serovar 1/2a;
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Ruenick C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Checouani F., Couve E., de Darvar A., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomics of Listeria species.";
RT Science 294:849-852(2001).
RL
CC -1- FUNCTION: The UVRABC repair system catalyzes the recognition and
CC processing of DNA lesions. UVR_C both incises the 5' and 3' sides
CC of the lesion. The N-terminal half is responsible for the 3'
CC incision and the C-terminal half is responsible for the 5'
CC incision (By similarity).
CC -1- SUBUNIT: Interacts with UVR_B in an incision complex (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the uvr_C family.
CC -1- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL591978; CAC9312.1; -.
CC PIR; A1229; A1229.
CC Listlist; LMO1234; -.
CC HAMAP; MF_00203; -.
CC InterPro; IPR003583; HHH 1.
CC InterPro; IPR001943; UVRB/C.
CC InterPro; IPR004791; UVR_C.
CC InterPro; IPR001162; UVR_C.
CC InterPro; IPR000305; UVR_C.
CC Pfam; PF01541; Excl_endo_N; 1.
CC Pfam; PF02151; UVR; 1.
CC ProDom; PD005870; UVR_C; 1.
CC SMART; SM00465; G1YC; 1.
CC SMART; SM00278; HhH1; 1.
CC TIGRfam; TIGR00194; uvr_C; 1.
CC PROSITE; PS50151; UVR; 1.
CC PROSITE; PS50164; UVR_C_1; 1.

```

DR PROSITE; PS50165; UVR\_C2; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KM DNA excision; Complete proteome.  
 FT DOMAIN 197 232 UVR.  
 SQ SEQUENCE 603 AA; 69315 MM; E87d742AF5F3325B CRC64;

Query Match 42.3%; Score 46.5; DB 1; Length 603;  
 Best Local Similarity 45.5%; Pred. No. 12;  
 Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

QY 1 VPSYSPARADSEVQDLIRDT 22  
 DB 42 VRSY-PSGTHSXTORLVOEIV 62

RESULT 10  
 UBI\_E RHILLO STANDARD; PRT; 258 AA.  
 AC 098GV1;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ubiqunone/menaquinone biosynthesis methyltransferase ubiE  
 DE (EC 2.1.1.-).  
 GN UBI\_E OR MLR3165.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 NCBI\_TaxID=381;  
 RX MEDLINE=22421331; PubMed=11214968;  
 RA KANEKO T., NAKAMURA Y., SATO S., ASAMIZU E., KATO T., SASAMOTO S.,  
 RA KISHIDA Y., KIKUYAMA K., ISHIKAWA A., KAWASHIMA K., KIMURA T.,  
 RA KOSHIZUKA Y., NAKAYAMA S., NAKAZAKI N., SHIMPO S., SUGIMOTO M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 CC -1- FUNCTION: Methyltransferase required for the conversion of  
 dimethylmenaquinone (DMK2) to menaquinone (MK2) and the  
 conversion of 2-polyprenyl-6-methoxy-1,4-benzoquinol (DDMQ2) to  
 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol (DMQ2) (By  
 similarity).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 2-polyprenyl-6-  
 methoxy-1,4-benzoquinol = S-adenosyl-L-homocysteine + 2-  
 polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + demethylmenaquinol =  
 S-adenosyl-L-homocysteine + menaquinol.  
 CC -1- PATHWAY: Menaquinone biosynthesis; last step.  
 CC -1- SIMILARITY: Ubiqunone biosynthesis.  
 CC -1- SIMILARITY: Belongs to the ubiE family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF003001; BAB50115.1; -.  
 DR HAMAP; MF 01813; -; 1.  
 DR InterPro; IPR001601; Methyltransf.  
 DR InterPro; IPR000051; SAM\_bnd.  
 DR InterPro; IPR004034; Ubi\_men\_Mettransf.  
 DR InterPro; IPR004033; UbiE/COG6\_Mettrf.  
 DR Pfam; PF01209; UbiE\_methyltransf. 1.  
 DR PROSITE; PS01183; UBI\_E 1; 1.  
 DR PROSITE; PS01184; UBI\_E 2; 1.  
 KW Menaquinone biosynthesis; Ubiqunone biosynthesis; Transferase;

KW Methyltransferase; Complete proteome.  
 SQ SEQUENCE 258 AA; 28291 MM; 9909682B726B2828 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 258;  
 Best Local Similarity 42.1%; Pred. No. 7.8;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 SYSPARADSEVQDLIRDT 21  
 DB 15 SYGPKRVEGKQSLVNDV 33

RESULT 11  
 GSCP\_XYLET STANDARD; PRT; 993 AA.  
 ID GSCP\_XYLET  
 AC 087DR1;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glycine decarboxylase (decarboxylating) (EC 1.4.4.2) (glycine  
 decarboxylase) (glycine cleavage system P-protein).  
 GN GSCP OR PD0620.  
 OS Xylella fastidiosa (strain Temecula / ATCC 700964).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 NCBI\_TaxID=183190;  
 RX MEDLINE=22421331; PubMed=12533478;  
 RA VAN SLUYVS M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,  
 RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,  
 RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,  
 RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorry H., Trai S.M.,  
 RA Carlier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,  
 RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,  
 RA Martino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.V.,  
 RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,  
 RA da Cunha A.F., Penille R.C., Ferro J.A., Fomighieri E.F., Kishi L.T.,  
 RA Leon S.G., Oliveira A.R., Rosa V.E. Jr., Saesaki F.T., Sena J.A.D.,  
 RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaro L.G.,  
 RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,  
 RA Kitajima J.P.;  
 RT "Comparative analyses of the complete genome sequences of Pierce's  
 disease and citrus variegated chlorosis strains of Xylella  
 fastidiosa.";  
 RL J. Bacteriol. 185:1018-1026(2003).  
 CC -1- FUNCTION: The glycine cleavage system catalyzes the degradation of  
 glycine. The P protein binds the alpha-amino group of glycine  
 through its pyridoxal phosphate cofactor; CO(2) is released and  
 the remaining methylamine moiety is then transferred to the  
 lipamide cofactor of the H protein (By similarity).  
 CC -1- CATALYTIC ACTIVITY: glycine + lipoylprotein = S-  
 aminomethylidihydroilipoylprotein + CO(2).  
 CC -1- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -1- SUBUNIT: The glycine cleavage system is composed of four proteins:  
 P, T, L and H (By similarity).  
 CC -1- SIMILARITY: Belongs to the gcvP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB012555; AAO28492.1; -.  
 DR HAMAP; MF 00711; -; 1.  
 DR InterPro; IPR003437; GDC-P.  
 DR Pfam; PF02347; GDC-P; 1.  
 DR Oxidoreductase; Pyridoxal phosphate; Complete proteome.  
 FT BINDING 715 715 PYRIDOXAL PHOSPHATE (By SIMILARITY).  
 SQ SEQUENCE 993 AA; 107646 MM; BCC04B29C442P43 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 993;  
Best Local Similarity 40.0%; Pred. No. 35;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 VPSYSPARAHSEVODLIR 20  
DB 969 IPVEAYKEKGSSEIOLIEE 988

## RESULT 12

ID POL3 DROME STANDARD; PRT; 1058 AA.  
AC P04323;  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Retrovirus-related Pol polyprotein from transposon 17.6 [Contains:  
Protease (EC 3.4.23.-); Reverse transcriptase (EC 2.7.7.49);  
Endonuclease].  
GN POL.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85061628; PubMed=6209583;  
RA Saigo K., Kuyimiyu W., Matsuo Y., Inouye S., Yoshioke K., Yuki S.,  
RT "Identification of the coding sequence for a reverse  
transcriptase-like enzyme in a transposable genetic element in  
Drosophila melanogaster."  
RL Nature 312:659-661(1984).  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
+ {DNA} (N).  
CC -1- MISCELLANEOUS: The open reading frame is located in a copia-like  
transposable element called 17.6.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X01472; CAA25702.1; -.  
DR PIR; A03971; GNF17.  
DR MEROPS; A02\_052; -.  
DR FlyBase; FBgn0014453; 17.6(pol.  
DR InterPro; IPR001969; Aspprotease\_AS.  
DR InterPro; IPR009007; Pept\_A\_acid.  
DR InterPro; IPR001995; Peptidase\_A2.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvt; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;  
KW Endonuclease; Transferrase; Polyprotein; Transposable element.  
FT ACT SITE 30 30  
SQ SEQUENCE 1058 AA; 122697 MW; C893F5CA7E1F091 CRC64;

## RESULT 13

SYE STAE

ID SYE STAE

STANDARD; PRT; 484 AA.

AC Q8CTJ3;

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)

DE (GLURS).

GN GLTX OR SE0290.

OS Staphylococcus epidermidis.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=1282;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 12228;

RX PubMed=12950922;

RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,

RA Qian Z.-H., Zhao G.-P., Gu D., Danchin A., Wen Y.-M.,

RT "Genome-based analysis of virulence genes in a non-biofilm-forming

RT Staphylococcus epidermidis strain (ATCC 12228)."

RL Mol. Microbiol. 49:1577-1593(2003).

CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +

CC diphosphate + L-glutamyl-tRNA(Glu).

CC -1- SUBUNIT: Monomer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE016744; AAC03887.1; -.  
DR HMAP; MF 00022; -1  
DR InterPro; IPR004527; GLTX\_bact.  
DR InterPro; IPR000924; GLU\_tRNA-synt\_1c.  
DR InterPro; IPR008925; tRNA-synt\_bnd.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR Pfam; PF00749; tRNA-synt\_1c; 1.  
DR PRINTS; PR00987; TRNASYNTGLU.  
DR TIGRFAMs; TIGR00464; gltx\_bact; 1.  
DR PROSITE; PS00178; AA\_tRNA\_LIGASE\_1; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
FT SITE 11 21  
FT SITE 252 256  
FT BINDING 255 255  
SQ SEQUENCE 484 AA; 56370 MW; 644A72F0C8B24FEB CRC64;

Query Match 40.5%; Score 44.5; DB 1; Length 484;  
Best Local Similarity 45.5%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 1 VPSYSPARAHSEVODLIR 19  
DB 193 VPTYNFAVAVDHYWQISDVIR 214

## RESULT 14

YH75\_ARCFU

ID YH75\_ARCFU

STANDARD; PRT; 330 AA.

AC Q28439;

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Hypothetical protein APL775.

```

GN AF1775.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klank H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.B.,
RA Ketchum K.A., Dodson R.J., Gilm M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kertavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kitzness B.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.M., D'Andrea K.P., Bowman C., Fujil C., Garland S.A.,
RA Maeson T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: Belongs to the ATZ/TRZ family.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF000980; AAB89475.1; -
DR PIR; F69471; F69471.
DR TIGR; AF1775; -
DR InterPro; IPR006680; Amidohydro_1.
DR Pfam; PF01979; Amidohydro_1.
DR Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 330 AA; 37069 MW; 525AD97F35A6FB1 CRC64;
QY Query Match 40.0%; Score 44; DB 1; Length 330;
Best Local Similarity 42.1%; Pred. No. 15;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
DB 3 SYSFARHDSVQDLIRDI 21
152 AYSSARDHDKLMEYREI 170
RESULT 15
ANPC HUMAN STANDARD; PRT; 541 AA.
ID ANPC_HUMAN STANDARD; PRT; 541 AA.
AC P17342;
DT 01-AUG-1990 (Rel. 15, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Atrial natriuretic peptide clearance precursor (ANP-C)
DE (ANP-C) (NPP-C) (Atrial natriuretic peptide C-type receptor).
GN NPP3 OR ANP-C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Kidney;
RX MEDLINE=9028735; PubMed=2162522;
RA Lowe D.G., Camerata T.R., Goeddel D.V.;
RT "cDNA sequence of the human atrial natriuretic peptide clearance
RT receptor."
RL Nucleic Acids Res. 18:3412-3412(1990).
RN [2]

```

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=90386556; PubMed=2169733;
RA Porter J.G., Arfsten A., Fuller F., Miller J.A., Gregory L.C.,
RA Lewicki J.A.;
RT "Isolation and functional expression of the human atrial natriuretic
RT peptide clearance receptor cDNA."
RL Biochem. Biophys. Res. Commun. 171:796-803(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Lens epithelium;
RA Rae U.L., Shepard A.R.;
RT "Human lens epithelial mRNA for atrial natriuretic peptide clearance
RT receptor."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE. DOES NOT HAVE
CC GUANYLATE CYCLASE ACTIVITY.
CC -1- SUBUNIT: Homodimer; disulfide-linked.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Bvent-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P17342-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P17342-2; Sequence=VSP 001812;
CC -1- MISCELLANEOUS: THERE SEEM TO BE AT LEAST THREE ANP RECEPTORS: TWO
CC WITH GUANYLATE CYCLASE ACTIVITY (ANP-A AND ANP-B) AND ONE (ANP-C)
CC WHICH IS PROBABLY RESPONSIBLE FOR THE CLEARANCE OF ANP FROM THE
CC CIRCULATION WITHOUT A ROLE IN SIGNAL TRANSDUCTION.
CC -1- SIMILARITY: TO ANP-A AND ANP-B RECEPTORS IN THEIR EXTRACELLULAR
CC AND TRANSMEMBRANE DOMAINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X52282; CAA36523.1; -
DR EMBL; M59305; AAB51734.1; -
DR EMBL; AF025996; AAB8801.1; -
DR PIR; S10150; OYHOCR.
DR PDB; 1UDN; 05-SEP-01.
DR Genew; HGNC:7945; NPP3.
DR MIM; 108962; -
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0001501; F:skeletal development; TAS.
DR InterPro; IPR001828; ANP_receptor.
DR InterPro; IPR001170; Ntpep_receptor.
DR Pfam; PF01094; ANP_receptor; 1.
DR PRINTS; PR00255; NATPEPTIDER.
DR PROSITE; PS00458; ANP_RECEPTORS; 1.
DR Receptor; Glycoprotein; Transmembrane; Signal; Alternative splicing;
KW 3D-structure.
KM SIGNAL 1 26
FT CHAIN 27 541
FT DOMAIN 27 481
FT TRANSMEM 482 504
FT DOMAIN 505 541
FT DISULFID 108 136
FT DISULFID 213 261
FT DISULFID 473 473
FT CARBOHYD 86 96
FT CARBOHYD 293 293
FT CARBOHYD 394 394
FT VARSPLIC 476 477
SQ SEQUENCE 541 AA; 59807 MW; 8A66415F7FD62B7 CRC64;
QY Query Match 40.0%; Score 44; DB 1; Length 541;
Best Local Similarity 38.9%; Pred. No. 26;

```



Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 4 YSPFARHSEVQDLINDI 21  
 ||| :|:|:|:  
 Db 234 YSPFETKQDLIEDIVRNI 251

RESULT 16  
 BACC\_BACLI STANDARD; PRT; 6359 AA.  
 AC 068008;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Bactracin synthetase 3 (BAC3) [includes: ATP-dependent isoleucine  
 adenylation (IleA) (isoleucine activase); ATP-dependent D-phenylalanine  
 adenylation (D-PheA) (D-phenylalanine activase); ATP-dependent histidine  
 adenylation (HisA) (histidine activase); ATP-dependent D-aspartate  
 adenylation (D-AspA) (D-aspartate activase); ATP-dependent asparagine  
 adenylation (AsnA) (asparagine activase); Aspartate racemase  
 (EC 5.1.1.13); Phenylalanine racemase (ATP hydrolyzing)  
 (EC 5.1.1.11)].  
 DE BACC.  
 GN Bacillus licheniformis.  
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=1402;  
 RX MEDLINE=98069193; PubMed=9427658;  
 RA Konz D., Kleus A., Schoergerdorfer K., Marahiel M.A.;  
 RT "The bacteracin biosynthesis operon of Bacillus licheniformis ATCC  
 10716: molecular characterization of three multi-modular peptide  
 synthetases.";  
 RT Chem. Biol. 4:927-937(1997).  
 CC -1 FUNCTION: INDUCES PEPTIDE SYNTHESIS, ACTIVATES AND INCORPORATES  
 CC FIVE AMINO ACIDS, FORMS A THIAZOLINE RING BETWEEN THE FIRST TWO  
 CC AMINO ACIDS AND INCORPORATES A D-GLUTAMINE IN THE FOURTH POSITION.  
 CC -1 CATALYTIC ACTIVITY: L-aspartate = D-aspartate +  
 CC -1 CATALYTIC ACTIVITY: ATP + L-phenylalanine = AMP + diphosphate + D-  
 CC phenylalanine.  
 CC -1 COFACTOR: Contains 5 covalently bound phosphopantetheines  
 CC (potential).  
 CC -1 PATHWAY: Cyclic peptide antibiotic bacitracin biosynthesis.  
 CC -1 SUBUNIT: LARGE MULTIMERIC COMPLEX OF BAI, BA2 AND BA3.  
 CC -1 DOMAIN: CONSISTS OF FIVE MODULES WITH TWO EPIMERIZATION DOMAINS IN  
 CC THE SECOND AND FOURTH MODULES, AND A POTATIVE C-TERMINAL  
 CC THIOESTERASE DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO  
 CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS  
 CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOALATION, CONDENSATION  
 CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND  
 CC N METHYLATION (OPTIONAL).  
 CC -1 MISCELLANEOUS: BACTRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC  
 CC DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST  
 CC ABUNDANT IS BACTRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT  
 CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-  
 CC GLU-ILE) WITH AN ISOLEUCINE-CYSTINE THIAZOLINE CONDENSATION  
 CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-  
 CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-  
 CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT  
 CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,  
 CC PHE-9, AND ASP-11).  
 CC -1 SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme  
 CC family.  
 CC -1 SIMILARITY: Contains 5 acyl carrier domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF007865; AAC06348.1; --  
 DR PIR: T31679; T31679.  
 DR HSSP: P14687; 1AMU.  
 DR InterPro: IPR000873; AMP-bind.  
 DR InterPro: IPR001242; Condensatn.  
 DR InterPro: IPR006163; Pp\_bind.  
 DR InterPro: IPR006162; Pantine S.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00501; AMP-binding; 5.  
 DR Pfam: PF00668; Condensation; 7.  
 DR Pfam: PF00550; Pp-binding; 5.  
 DR Pfam: PF00975; Thioesterase; 1.  
 DR PRINTS: PR00154; AMPBINDING.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 4.  
 DR PROSITE: PS00455; AMP BINDING; 5.  
 DR PROSITE: PS00705; ACP DOMAIN; 5.  
 DR PROSITE: PS00705; ACP DOMAIN; 5.  
 KW Ligase; Isomerase; Hydrolyase; Multifunctional enzyme; Repeat.  
 KW Phosphopantetheine; Multifunctional enzyme; Repeat.  
 FT REPEAT 461 1034 DOMAIN 1 (ISOLEUCINE-ACTIVATING).  
 FT REPEAT 1517 2064 DOMAIN 2 (D-PHENYLALANINE-ACTIVATING).  
 FT REPEAT 2999 3570 DOMAIN 3 (HISTIDINE-ACTIVATING).  
 FT REPEAT 4047 4612 DOMAIN 4 (D-ASPARTIC ACID-ACTIVATING).  
 FT REPEAT 5549 6129 DOMAIN 5 (ASPARAGINE-ACTIVATING).  
 FT DOMAIN 966 1033 ACTL CARRIER (ACP) 1.  
 FT DOMAIN 1998 2064 ACTL CARRIER (ACP) 2.  
 FT DOMAIN 3502 3569 ACTL CARRIER (ACP) 3.  
 FT DOMAIN 4544 4610 ACTL CARRIER (ACP) 4.  
 FT DOMAIN 6052 6119 ACTL CARRIER (ACP) 5.  
 FT BINDING 996 996 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT BINDING 2028 2028 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT BINDING 3532 3532 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT BINDING 4574 4574 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT BINDING 6082 6082 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 SQ SEQUENCE 6359 AA; 722923 MW; 82A273C54625074 CRC64;

QY 1 YSPFARHSEVQDLINDI 22  
 ||| :|:|:|:  
 Db 3186 VPSFSEF---DSVEIDFTLLI 3203

Query Match 40.0%; Score 44; DB 1; Length 6359;  
 Best Local Similarity 45.5%; Pred. No. 4; 1e+02;  
 Matches 10; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

RESULT 17  
 SYE\_STAM STANDARD; PRT; 484 AA.  
 ID SYE\_STAM  
 AC Q99W75;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE [Glutamy]l-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)  
 DE (GURS).  
 GN GLTX OR SAV0528 OR SA0486 OR MW0483;  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
 OS Staphylococcus aureus (strain N315), and  
 OS Staphylococcus aureus (strain MW2).  
 CC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CC NCBI\_TaxID=158878, 158879, 196620;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=Mu50 / ATCC 700699, and N315;  
 CC MEDLINE=21311952; PubMed=11418146;  
 CC Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 CC Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,  
 CC Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 CC Mitutani O-I Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaio C.,  
 CC Sekimizu K., Hiraoka S., Kuhara S., Goto S., Yabuzaki J.,  
 CC Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 CC Hattori M., Ogasawara N., Hayashi H., Hiratake K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus.";



```

RL Lancel 357:1225-1240(2001).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-MM2:
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda K., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwano N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.,
RT "Genome and virulence determinants of high virulence community-
acquired MRSA."
RL Lancel 359:1819-1827(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP003359; BAB56690.1; -
CC EMBL: AP003130; BAB41716.1; -
CC EMBL: AP004823; BAB94348.1; -
CC PIR: A89820; A89820.
CC HSSP: P27000; 1GLX.
CC DR HSSP: P27000; 1GLX.
CC DR SWISS-2DPAGE: Q9N75; STRAN.
CC DR HAMAP: MF_00022; -; 1.
CC DR InterPro: IPR004527; GlxL bact.
CC DR InterPro: IPR000924; Glu tRNA-synt 1c.
CC DR InterPro: IPR008925; tRNA-synt bind.
CC DR InterPro: IPR001412; tRNA-synt 1.
CC Pfam: PF00749; tRNA-synt 1c; 1.
CC DR PRINTS: PR00987; TRNASYNTHGLU.
CC DR TIGRFAMs: TIGR00464; glxL bact; 1.
CC DR PROSITE: PS00178; AA tRNA-LIGASE 1; 1.
CC DR AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC KW Complete proteome.
CC FT SITE 11 21 "HIGH" REGION.
CC FT SITE 252 256 "KMSKS" REGION.
CC FT BINDING 255 255 ATP (By SIMILARITY).
CC FT SEQUENCE 484 AA; 56288 MW; 4CBASFF08DA23BEFA CRC64;
SQ
Query Match 39.5%; Score 43.5; DB 1; Length 484;
Best Local Similarity 40.9%; Pred. No. 27;
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;
QY 1 VPSPFARA---HSEVQDLIR 19
Db 193 IPTYNFAVALDDHMQSDVIR 214

```

```

RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Deroff B.S., Estrem S.T., Fultz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass U.S., Kioja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Letkowitz E.J., Lu J., Matsushima P.,
RA McHenry S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Taekman S.R., Rostock P.R. Jr., Skarrud P.L.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE008552; AAL00683.1; -
CC PIR: F98106; F98106.
CC DR HAMAP: MF_00022; -; 1.
CC DR InterPro: IPR004527; GlxL bact.
CC DR InterPro: IPR000924; Glu tRNA-synt 1c.
CC DR InterPro: IPR008925; tRNA-synt bind.
CC DR InterPro: IPR001412; tRNA-synt 1.
CC Pfam: PF00749; tRNA-synt 1c; 1.
CC DR PRINTS: PR00987; TRNASYNTHGLU.
CC DR TIGRFAMs: TIGR00464; glxL bact; 1.
CC DR PROSITE: PS00178; AA tRNA-LIGASE 1; 1.
CC DR AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC KW Complete proteome.
CC FT SITE 11 21 "HIGH" REGION.
CC FT SITE 255 259 "KMSKS" REGION.
CC FT BINDING 258 258 ATP (By SIMILARITY).
CC FT SEQUENCE 486 AA; 55912 MW; F7315B21CB2381D CRC64;
SQ
Query Match 38.6%; Score 42.5; DB 1; Length 486;
Best Local Similarity 42.9%; Pred. No. 40;
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;
QY 2 PYSFARA---HSEVQDLIR 19
Db 196 PTYNFAVALDDHMQSDVIR 216

```

```

RESULT 18
SYE STRR6
ID ID SYE STRR6 STANDARD; PRT; 486 AA.
AC Q8CRN5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (GluRS).
DE GN GTX OR SPR1881.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;

```

```

RESULT 19
HXTD YEAST
ID HXTD YEAST STANDARD; PRT; 540 AA.
AC P42833;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hexose transporter HXT14.
GN HXT14 OR HXT9 OR YNLJ18C OR N0345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=8288C / FY1676;
RX MEDLINE=96076632; PubMed=7502583;
RA Maffei M., Nicoud J.-M., Levesque H., Gaillardin C.;
RT "Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV
RT identifies six known genes, a new member of the hexose transporter
RT family and ten new open reading frames."

```

```

RL Yeast 11:1077-1085(1995).
CC -1- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the sugar transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46259; -; NOT ANNOTATED_CDS.
DR EMBL; Z71595; CA96250.1; -.
DR PIR; S63299; S63299.
DR GerMOnline; I43324; -.
DR SGD; S0005262; HXT14.
DR GO; GO:0005354; F:galactose transporter activity; IGI.
DR GO; GO:0008645; P:hexose transport; IGI.
DR InterPro; IPR007114; MPS.
DR InterPro; IPR005828; Sub transporter.
DR InterPro; IPR005829; Sug transporter.
DR InterPro; IPR003653; Sugar_transpc.
DR Pfam; PF00083; sugar_tr; 1.
DR PRINTS; PR00171; SUGTRANSFPORT.
DR TIGRPFAMS; TIGR00879; SP; 1.
DR PROSITE; PS00850; MPS; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR Repeat; Transmembrane; Sugar transport; Transport.
KT DOMAIN 1 56
FT TRANSEM 57 76
FT DOMAIN 77 119
FT TRANSEM 120 140
FT DOMAIN 141 146
FT TRANSEM 147 167
FT DOMAIN 168 177
FT TRANSEM 178 198
FT DOMAIN 199 204
FT TRANSEM 205 225
FT DOMAIN 226 243
FT TRANSEM 244 264
FT DOMAIN 265 357
FT TRANSEM 358 374
FT DOMAIN 375 380
FT TRANSEM 381 398
FT DOMAIN 399 405
FT TRANSEM 406 426
FT DOMAIN 427 440
FT TRANSEM 441 461
FT DOMAIN 462 478
FT TRANSEM 479 499
FT DOMAIN 500 500
FT TRANSEM 501 521
FT DOMAIN 522 540
SQ SEQUENCE 540 AA; 60978 MW; 91A6BBA27099EEF8 CRC64;

Query Match 38.6%; Score 42.5; DB 1; Length 540;
Best Local Similarity 48.3%; Pred. No. 45;
Matches 14; Conservative 3; Mismatches 5; Indels 7; Gaps 2;

CY 1 VPS--YSPAR-----AHDSVQDLIRDI 22
DB 277 VPSAKTSPKAWNGIPATDSWTERFDLL 305

RESULT 20
MAD2_YEAST STANDARD; PRT; 196 AA.
AC P40958.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mitotic spindle checkpoint component MAD2 (Mitotic MAD2 protein).
GN MAD2 OR YJL030W OR J1256.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Li R., Havel C., Watson J.A., Murray A.W.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Pohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=91330300; PubMed=1651172;
RA Li R., Murray A.W.;
RT "Feedback control of mitosis in budding yeast.";
RL Cell 66:519-531(1991).
RN [4]
RP INTERACTIONS.
RX MEDLINE=98128031; PubMed=9461437;
RA Hwang L.H., Lau L.F., Smith D.L., Mistrot C.A., Hardwick K.G.,
RA Hwang E.S., Amon A., Murray A.W.;
RT "Budding yeast Cdc20: a target of the spindle checkpoint.";
RL Science 279:1041-1044(1998).
CC -1- FUNCTION: Feedback control that prevents cells with incompletely
CC assembled spindles from leaving mitosis.
CC -1- SUBUNIT: The spindle checkpoint complex is composed of MAD1, MAD2
CC and MAD3. It interacts with CDC20.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE MAD2 FAMILY.
CC -1- SIMILARITY: Contains 1 HORNA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14132; AAA21385.1; -.
DR EMBL; Z49305; CA89321.1; -.
DR PIR; S48302; S48302.
DR HSSP; Q13257; IDUJ.
DR GerMOnline; I41645; -.
DR SGD; S0003567; MAD2.
DR GO; GO:0000778; C:condensed nuclear chromosome kinetochore; IDA.
DR GO; GO:0005643; C:nuclear pore; IDA.
DR GO; GO:0007094; P:mitotic spindle checkpoint; IGI.
DR InterPro; IPR003511; DNABind_HORNA.
DR Pfam; PF02301; HORNA; 1.
DR PROSITE; PS0815; HORNA; 1.
DR Cell cycle; Mitosis; Nuclear protein.
KT DOMAIN 8 192
FT DOMAIN 192 2284 MW; EFES9916C5720644 CRC64;
SQ SEQUENCE 196 AA; 22284 MW;

Query Match 38.2%; Score 42; DB 1; Length 196;
Best Local Similarity 46.7%; Pred. No. 17;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

CY 8 RAHDSVQDLIRDI 22
DB 49 KTHDDELKDYIRKIL 63

RESULT 21
YIBQ_HAEIN STANDARD; PRT; 280 AA.
AC P44863.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein HI0755 precursor.
GN HI0755
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Rd / KM20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RX Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
RA Utecherback T.R., Hanna M.C., Nguyen D.T., Gaudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Ghem C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.;
RL Science 269:496-512(1995).

[2]
RN IDENTIFICATION BY MASS SPECTROMETRY.
RP MEDLINE=20137488; PubMed=10675023;
RX Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.,
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).

[3]
RN -1- SIMILARITY: STRONG, TO E.COLI Y1BQ.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32759; AAC22414.1; ALT_INIT.
CC DR TIGR; HI0755;
CC DR InterPro; IPR006837; DUF610.
CC DR Pfam; PF04748; div_psaacdeacet; 1.
CC KW Signal; Complete proteome.
CC FT SIGNAL 1 26 POTENTIAL.
CC FT CHAIN 27 280 PROTEIN HI0755.
CC SQ SEQUENCE 280 AA; 30740 MW; C739E02161B53B8 CRC64;

Query Match 38.2%; Score 42; DB 1; Length 280;
Best Local Similarity 36.4%; Pred. No. 26;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPSPARADSEVODIIRDI 22
DB 59 IPAPVBARNOEAKSGRDIL 80

RESULT 22
TF3A_YEAST STANDARD; PRT; 429 AA.
AC P39933;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor IIA (TFIIIA).
GN TF3C OR PZFI OR TFIIIA OR YPR186C OR P9677.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147684; PubMed=1737784;
RA Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA Segall J.;
RT "The deduced sequence of the transcription factor TFIIIA from
RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT TFIIIA.";
RL J. Biol. Chem. 267:3282-3288(1992).

[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=92237295; PubMed=1570325;
RX Wojcik N.A., Young R.A.;
RT "Genes encoding transcription factor IIA and the RNA polymerase
RT common subunit RPB6 are divergently transcribed in Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).

[3]
RN SEQUENCE FROM N.A.
RP STRAIN=S288c / AB972;
RC MEDLINE=97313271; PubMed=9169875;
RX Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Bortstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duestenhoef A.,
RA Duncan M., Floeth W., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunkeler-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
RA Komp C., Kurti O., Lashari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marthe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafert M., Scharte M.,
RA Scherrens B., Schramm S., Schroeder M., Slicu A.M., Tettelin H.,
RA Uristarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambolt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Han J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).

[4]
RN -1- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA'S.
CC -----
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60611; AAB08014.1;
CC DR EMBL; M90638; -; NOT ANNOTATED_CDS.
CC DR EMBL; U25841; AAB64615.1;
CC DR PIR; S20050; S20050.
CC DR Germonline; 144451;
CC DR TRANSFAC; T03530;
CC DR SGD; S0006390; PZFI.
CC DR GO; GO:0005667; C:transcription factor complex; IPI.
CC DR GO; GO:0003709; F:RNA polymerase III transcription factor act. .; IPI.
CC DR GO; GO:0006384; P:transcription initiation from Pol III promoter; IPI.
CC DR InterPro; IPR007087; Znf_C2H2.
CC DR Pfam; PF00096; ZF-C2H2; 9.
CC DR SMART; SM00355; ZNF_C2H2; 9.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
CC DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
CC KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
CC RNA-binding; Repeat; Nuclear protein.
CC FT DOMAIN 24 43 SER-RICH.
CC FT ZN_FING 49 74 C2H2-TYPE.
CC FT ZN_FING 80 102 C2H2-TYPE.

```

FT ZN PING 108 130 C2H2-TYPE.  
 FT ZN PING 134 159 C2H2-TYPE.  
 FT ZN PING 163 186 C2H2-TYPE.  
 FT ZN PING 194 219 C2H2-TYPE.  
 FT ZN PING 222 244 C2H2-TYPE.  
 FT ZN PING 253 277 C2H2-TYPE.  
 FT ZN PING 321 327 ARG/LYS-RICH (BASIC).  
 FT ZN PING 365 389 C2H2-TYPE.  
 SQ SEQUENCE 429 AA; 50027 MW; 209B1EDEBA20422D9 CRC64;

Query Match 38.2%; Score 42; DB 1; Length 429;  
 Best Local Similarity 24.4%; Pred. No. 42;  
 Matches 10; Conservative 5; Mismatches 4; Indels 22; Gaps 1;

QY 2 PYSYFARAD-----SEVODLRD 20  
 DB 259 PMSFRRKDLTHYGSHTEDDIPLELYKXISDIOQLVOD 299

RESULT 23  
 ID RPOC WEIHE STANDARD; PRT; 1046 AA.  
 AC P96A177;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase beta' chain) (RNA polymerase beta' subunit) (Fragment).  
 GN RPOC.  
 OS Weissella helvetica.  
 OC Bacteria; Firmicutes; Lactobacillales; Weissella.  
 OX NCBI\_TaxID=46256;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCFB 2973;  
 RX MEDLINE=97016803; PubMed=8863429;  
 RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;  
 RT "Analysis of the beta' subunit of DNA-dependent RNA polymerase does not support the hypothesis inferred from 16S rRNA analysis that RT encodes a (formerly Leuconostoc oenos) is a tachyelic (fast-evolving) bacterium."  
 RT Int. J. Syst. Bacteriol. 46:1004-1009 (1996).  
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).  
 CC -1- SUBUNIT: The enzyme consists of the sigma chain and the core enzyme which is composed of 2 alpha chains, 1 beta chain, and 1 beta chain.  
 CC -1- SIMILARITY: Belongs to the RNA polymerase beta' chain family.  
 CC -1- SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X96470; CA65322.1; --  
 DR HSSP; Q9KTM6; 1HOM.  
 DR InterPro; IPR007072; RNA\_pol\_A.  
 DR InterPro; IPR007080; RNA\_pol\_Rpbl\_1.  
 DR InterPro; IPR007066; RNA\_pol\_Rpbl\_3.  
 DR InterPro; IPR007083; RNA\_pol\_Rpbl\_4.  
 DR InterPro; IPR007081; RNA\_pol\_Rpbl\_5.  
 DR InterPro; IPR006592; RNA\_pol\_N.  
 DR Pfam; PF04997; RNA\_pol\_Rpbl\_1; 1.  
 DR Pfam; PF00623; RNA\_pol\_Rpbl\_2; 1.  
 DR Pfam; PF04983; RNA\_pol\_Rpbl\_3; 1.  
 DR Pfam; PF05000; RNA\_pol\_Rpbl\_4; 1.  
 DR Pfam; PF04998; RNA\_pol\_Rpbl\_5; 1.

DR SMART: SM00663; RPOLA\_N; 1.  
 KM Transferase; DNA-directed RNA polymerase; Transcription.  
 FT NON\_TER 1 1  
 FT NON\_TER 1046 1046  
 SQ SEQUENCE 1046 AA; 117107 MW; D54C62C26A7F1696 CRC64;

Query Match 38.2%; Score 42; DB 1; Length 1046;  
 Best Local Similarity 46.2%; Pred. No. 1; Le+02;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 10 HSEVODLRDII 22  
 DB 327 HDEVDVMDVLEVDY 339

RESULT 24  
 ID PRAX RAT STANDARD; PRT; 1383 AA.  
 AC O63425;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Periaxin.  
 GN PRX.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=Miscar.  
 RX MEDLINE=94206531; PubMed=8155317;  
 RA Gillespie C.S., Sherman D.L., Blair G.E., Brophy P.J.;  
 RT "Periaxin, a novel protein of myelinating Schwann cells with a possible role in axonal ensheathment."  
 RT Neuron 12:497-508 (1994).  
 RL [2]  
 RP REVISIONS TO 1364-1371.  
 RA Brophy P.J.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=98157980; PubMed=9488714;  
 RA Dyttrich L., Sherman D.L., Gillespie C.S., Brophy P.J.;  
 RT "Two PDZ domain proteins encoded by the murine periaxin gene are the result of alternative intron retention and are differentially targeted in Schwann cells."  
 RT J. Biol. Chem. 273:5794-5800 (1998).  
 RL -1- FUNCTION: Seems to be required for maintenance of peripheral nerve myelin sheath. May have a role in axon-glia interactions, possibly by interacting with the cytoplasmic domains of integral membrane proteins such as myelin-associated glycoprotein in the periaxonal regions of the Schwann cell plasma membrane. May have a role in the early phases of myelin deposition.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing: Named isoforms=2;  
 CC Name=1; Synonyms=L-periaxin;  
 CC IsoId=Q63425-1; Sequence=Displayed;  
 CC Name=2; Synonyms=S-periaxin;  
 CC IsoId=Q63425-2; Sequence=VSP\_004368; VSP\_004369;  
 CC -1- TISSUE SPECIFICITY: NEURONES, SCHWANN CELL-SPECIFIC.  
 CC -1- DEVELOPMENTAL STAGE: MRNA AND PROTEIN LEVELS PEAK IN THE SCIENTIC NERVE BETWEEN POSTNATAL DAYS 8 AND 20; THEREAFTER THEY DECLINE PRECIPITOUSLY.  
 CC -1- DOMAIN: Has a remarkable domain of repetitive pentameric units sometimes followed by a tripeptide spacer. It may separate two functional basic and acidic domains.  
 CC -1- DOMAIN: The Arg/Lys-rich basic domain functions as a tripartite nuclear localization signal.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- SIMILARITY: Belongs to the periaxin family.  
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.



```

RESULT 26
Y326 MYCGB STANDARD; PRT; 295 AA.
AC P47568;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0230 protein M3326.
GN M3326.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
MEDLINE=96026346; PubMed=7569993;
RA Frazer C.M., Gecayane J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fleischmann J.L., Weidman J.F., Small K.V., Sandhu M., Frittmann J.L.,
RA Nguyen D.T., Uettersack T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0230 (DEGV) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39714; AAC71550.1; --
CC PIR: A64236; A64236.
CC TIGR: M3326; --
CC InterPro: IPR003797; DegV.
CC Pfam: PF02645; DUF194; 1.
CC TIGRfam: TIGR00762; DegV, 1.
CC DR Hypothetical protein; Complete proteome.
CC KW SEQUENCE 295 AA; 33413 MW; 04610881C0F841EE CRC64;
SQ
Query Match 37.3%; Score 41; DB 1; Length 295;
Best Local Similarity 36.8%; Pred. No. 40;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 4 YSFARADSEVQDLIRDI 22
DB 239 YSFCKNYANEIKITDPI 257

```

```

RL J. Neurosci. 10:412-419(1990).
RN
RP SEQUENCE FROM N.A.
RA Kelleet E.;
RL Thesis (1992), University of Sussex, U.K.
RN
RP PARTIAL SEQUENCE FROM N.A.
RC TISSUE=CNS;
RX MEDLINE=92185507; PubMed=1347559;
RA Saunders S.E., Kelleet E., Bright K., Benjamin P.R., Burke J.F.;
RT "Cell-specific alternative RNA splicing of an FMRamide gene
RT transcript in the brain.";
RL J. Neurosci. 12:1033-1039(1992).
[4]
RP SEQUENCE OF 82-103 (PN).
RC TISSUE=CNS;
RX MEDLINE=94108633; PubMed=7904219;
RA Santana N., Li K.W., Bright K.E., Yeoman M., Geraets W.P.M.,
RA Benjamin P.R., Burke J.F.;
RT "Processing of the FMRamide precursor protein in the snail Lymnaea
RT stagnalis: characterization and neuronal localization of a novel
RT peptide, 'SERPLY'.";
RL J. Neurosci. 5:1003-1016(1993).
CC -1- FUNCTION: FMRamide induces contractions in visceral and somatic
CC musculature as well as in the heart. May play a role as
CC cotransmitters or modulators in a number of significant neuronal
CC systems.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=isoform 1 and isoform 2 only share the N-terminal signal
CC sequence.
CC Name=1; Synonyms=FMRamide;
CC IsoId=P19802-1; Sequence=displayed;
CC Name=2; Synonyms=FMRamide-related;
CC IsoId=P42565-1; Sequence=external;
CC Name=3;
CC IsoId=P19802-2; Sequence=YSP_001564;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Expressed in 280 cells of the CNS including
CC the BGP heart excitatory motoneurons.
CC -1- SIMILARITY: Belongs to the FARP (FMRamide related peptide)
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M37629; AAA63280.1; --
CC DR EMBL: M87479; -; NOT ANNOTATED_CDS.
CC DR EMBL: S18686; AAB21767.1; --
CC DR EMBL: S94982; AAB21764.1; --
CC PIR: A37016; A37016.
CC InterPro: IPR002544; FARP.
CC Pfam: PF01581; FARP; 13.
CC Annotation: Cleavage on pair of basic residues; Repeat; Signal;
KW Neuropeptide; Alternative splicing.
FT SIGNAL 1 35
FT PROPEP 36 37
FT PEPTIDE 40 43 FLRF-AMIDE 1.
FT PROPEP 46 56
FT PEPTIDE 59 63 QFYRI-AMIDE.
FT PROPEP 66 73
FT PEPTIDE 76 79
FT PEPTIDE 82 103 FLRF-AMIDE 2.
FT PROPEP 108 149 PN (SERPLY).
FT PEPTIDE 152 155 FMRF-AMIDE 1.
FT PROPEP 158 163
FT PEPTIDE 166 169 FMRF-AMIDE 2.
FT PEPTIDE 173 176 FMRF-AMIDE 3.

```









QY 8 RAHSEVQDLIRDI 21  
| : | | | |  
Db 272 RIFRAAQDLFRDI 285

Search completed: May 4, 2004, 09:10:38  
Job time : 11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:05:56 ; Search time 33 Seconds  
(without alignments)  
210.345 Million cell updates/sec

Title: US-09-290-049a-17  
Perfect score: 110  
Sequence: 1 VPSYSPARAHSEVODLRDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

SPREMBL.25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	98.2	1590	2	Q59983 streptococc
2	108	98.2	1590	2	Q55263 streptococc
3	83	75.5	1290	2	Q48756 streptococc
4	80	72.7	1330	2	Q84CM4 leuconostoc
5	80	72.7	1477	2	Q91466 leuconostoc
6	80	72.7	1508	2	Q9EZH5 leuconostoc
7	80	72.7	1508	2	Q52224 leuconostoc
8	79	71.8	1016	2	Q91CJ7 leuconostoc
9	79	71.8	1527	2	Q9ZAR4 leuconostoc
10	79	71.8	1527	2	Q8KRE1 streptococc
11	78	70.9	1554	2	Q8KZL5 streptococc
12	74	67.3	1512	2	Q9WKJ5 streptococc
13	73	66.4	2835	2	Q8G9Q2 leuconostoc
14	70	63.6	1338	2	Q9WKJ4 streptococc
15	70	63.6	1575	2	Q91CH3 streptococc
16	70	63.6	1577	2	Q54178 streptococc

17	70	63.6	1599	2	Q00599 streptococc
18	68	61.8	1449	2	Q68542 streptococc
19	68	61.8	1449	2	Q55264 streptococc
20	68	61.8	2057	2	Q9RB05 streptococc
21	65	59.1	1518	2	Q00600 streptococc
22	62	56.4	1577	2	Q55265 streptococc
23	62	56.4	93	2	Q9ZIX9 streptococc
24	50	45.5	1554	3	Q8J0Z6 cryptococcu
25	50	45.5	1567	3	Q8J0Z6 cryptococcu
26	49	44.5	108	16	P74028 streptococc
27	49	44.5	303	16	Q8A338 streptococc
28	47	42.7	51	16	Q9RNP7 streptococc
29	47	42.7	462	16	Q983M9 streptococc
30	47	42.7	597	16	Q9PR58 streptococc
31	47	42.7	1516	10	Q7XW87 streptococc
32	46.5	42.3	2470	12	Q8U2G0 streptococc
33	46.5	42.3	2470	12	Q7YU22 streptococc
34	46	41.8	501	5	Q964R1 streptococc
35	46	41.8	1532	10	Q7XWU3 streptococc
36	45	40.9	209	16	Q82TR1 streptococc
37	45	40.9	282	16	Q7VLS4 streptococc
38	45	40.9	449	16	Q8R6U3 streptococc
39	45	40.9	533	2	Q8RUD0 streptococc
40	45	40.9	881	16	Q8G5W2 streptococc
41	45	40.9	2627	4	Q99973 streptococc
42	44.5	40.5	104	2	Q9X771 streptococc
43	44	40.0	200	16	Q8XUV7 streptococc
44	44	40.0	367	4	Q9H769 streptococc
45	44	40.0	480	5	Q95MU6 streptococc
46	44	40.0	492	2	Q8G6M5 streptococc
47	44	40.0	506	4	Q9N6L5 streptococc
48	44	40.0	566	12	Q8QPY3 streptococc
49	44	40.0	588	10	Q7XWU7 streptococc
50	44	40.0	604	16	Q8RUM3 streptococc
51	44	40.0	620	10	Q7XW15 streptococc
52	44	40.0	651	5	Q20710 streptococc
53	44	40.0	743	4	Q9NVJ7 streptococc
54	44	40.0	743	4	Q9NVJ7 streptococc
55	44	40.0	743	4	Q9BUN0 streptococc
56	44	40.0	779	4	Q9H3P4 streptococc
57	44	40.0	806	10	Q7XS18 streptococc
58	44	40.0	999	10	Q7XUY0 streptococc
59	44	40.0	1078	10	Q8LIX9 streptococc
60	44	40.0	1094	10	Q7XEP7 streptococc
61	44	40.0	1167	10	Q8S837 streptococc
62	44	40.0	1167	10	Q7XEX8 streptococc
63	44	40.0	1342	10	Q8S692 streptococc
64	44	40.0	1379	10	Q84RY1 streptococc
65	44	40.0	1406	10	Q8LWV8 streptococc
66	44	40.0	1412	10	Q7XNV8 streptococc
67	44	40.0	1416	10	Q8W5E8 streptococc
68	44	40.0	1455	10	Q8W089 streptococc
69	44	40.0	1470	10	Q7XLV2 streptococc
70	44	40.0	1473	10	Q8SB03 streptococc
71	44	40.0	1473	10	Q7XWPS streptococc
72	44	40.0	1473	10	Q7XN96 streptococc
73	44	40.0	1473	10	Q7XG94 streptococc
74	44	40.0	1475	10	Q8LHE4 streptococc
75	44	40.0	1484	10	Q7XW70 streptococc
76	44	40.0	1522	10	Q7XEP0 streptococc
77	44	40.0	1528	10	Q7XRP3 streptococc
78	44	40.0	1594	10	Q8S782 streptococc
79	44	40.0	1594	10	Q7XFS6 streptococc
80	44	40.0	1594	10	Q7XFS6 streptococc
81	44	40.0	1664	10	Q7XS20 streptococc
82	44	40.0	1666	10	Q7XRP5 streptococc
83	44	40.0	1679	10	Q7XRP8 streptococc
84	44	40.0	1683	10	Q8LMA6 streptococc
85	44	40.0	1722	10	Q8S644 streptococc
86	44	40.0	1729	10	Q8LIZ0 streptococc
87	44	40.0	1729	10	Q8LIZ0 streptococc
88	44	40.0	1729	10	Q8LIZ0 streptococc
89	44	40.0	1729	10	Q7XGP6 streptococc

90 44 40.0 1770 10 Q851D3  
 91 44 40.0 1779 10 Q7XX75  
 92 44 40.0 1963 10 Q7XW79  
 93 44 40.0 2118 10 Q7XS86  
 94 44 40.0 2157 10 Q9AY85  
 95 44 40.0 2157 10 Q7X978  
 96 44 40.0 2219 4 Q9C0A3  
 97 44 40.0 2296 4 Q9Y3S1  
 98 43.5 39.5 504 2 Q83ZD6  
 99 43 39.1 147 10 Q7XIX0  
 100 43 39.1 176 16 Q9K051

## ALIGNMENTS

## RESULT 1

Q59983 PRELIMINARY; PRT; 1590 AA.

AC Q59983: 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5).

GN GTFI.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_Taxid=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OM2176;

RX MEDLINE=94146405; PubMed=8312602;

RT Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.,

RL "DNA sequence of the glucosyltransferase gene of serotype d

EMBL; D13858; BAA02976.1; -

DR PIR; A39841; A39841.

DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.

DR GO; GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW binding\_1; 13.

DR Glycosyltransferase; signal; Transferase.

KW SIGNAL

FT CHAIN 39 1590 GLUCOSYLTTRANSFERASE-I.

SO SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B08 CRC64;

Query Match 98.2%; Score 108; DB 2; Length 1590;

Best Local Similarity 95.5%; Pred. No. 3.6e-08;

Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYFARADSEVQDIIRDI 22

DB 548 VPSYFARADSEVQDIIRDI 569

RESULT 2

Q55263 PRELIMINARY; PRT; 1590 AA.

AC Q55263: 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE GTF-I.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_Taxid=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33478;

RX "DNA and amino-acid sequences of water-insoluble-glucan synthetase

RT produced from Streptococcus sobrinus ATCC 33478."

RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).

DR EMBL; D63570; BAA09792.1; -

DR PIR; A39841; A39841.

DR GO; GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW binding\_1; 13.

DR Glycosyltransferase; signal; Transferase.

KW SIGNAL

FT CHAIN 39 1590 GLUCOSYLTTRANSFERASE-I.

SO SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B08 CRC64;

Query Match 98.2%; Score 108; DB 2; Length 1590;

Best Local Similarity 95.5%; Pred. No. 3.6e-08;

Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYFARADSEVQDIIRDI 22

DB 548 VPSYFARADSEVQDIIRDI 569

RESULT 3

Q48756 PRELIMINARY; PRT; 1290 AA.

AC Q48756: 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Dextranucrase.

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.

OX NCBI\_Taxid=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B1299.

RX MEDLINE=97136686; PubMed=8982063;

RT Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsen P.,

RL "Cloning and sequencing of a gene coding for a novel dextranucrase

RT from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-

RT 6) and alpha (1-3) linkages."

DR Gene 182:23-32(1996).

DR EMBL; U8181; AAB40875.1; -

DR PIR; JCS473; JCS473.

DR GO; GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW binding\_1; 9.

DR Glycosyltransferase; signal; Transferase.

KW SIGNAL

FT CHAIN 39 1590 GLUCOSYLTTRANSFERASE-I.

SO SEQUENCE 1290 AA; 145590 MW; 3555C2B96B749FMA CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1290;

Best Local Similarity 76.2%; Pred. No. 0.00029;

Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PYSYFARADSEVQDIIRDI 22

DB 388 PYSYFARADSEVQDIIRDI 408

RESULT 4

Q84CN4 PRELIMINARY; PRT; 1330 AA.

AC Q84CN4: 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Dextranucrase Dextr (EC 2.4.1.5).

GN DSNR.

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.

OC NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1501;  
 RA Kim C.H., Moon J.O., Jang E.K.;  
 RT "Gene encoding a dextranucrase (Dxr) in Leuconostoc mesenteroides  
 NRRL B-1501."  
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY142210; AAC8835.1; -  
 DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF02324; Glyco\_hydro\_1; 11.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1330 AA; 148863 MW; D945CB836C75797 CRC64;  
 Query Match 72.7%; Score 80; DB 2; Length 1330;  
 Best Local Similarity 68.2%; Pred. No. 0.00092; Mismatches 4; Indels 0; Gaps 0;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VPSYSPARADSEVDLIRDI 22  
 DB 456 IPNYSFVRADSEVQTIAQII 477  
 RESULT 5  
 Q9L466 PRELIMINARY; PRT; 1477 AA.  
 AC Q9L466;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Dextranucrase (EC 2.4.1.5).  
 GN DSRB.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1355;  
 RA Arguello-Morales M.A., Renaud-Simeon M., Pizut S., Sargabal P.,  
 RA Willemot R.M., Monan P.;  
 RT "Sequence analysis of the gene encoding altermannucrase, a sucrose  
 RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355."  
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ250172; CAB76565.1; -  
 DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding; 1; 13.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1477 AA; 164887 MW; E6F5710DEDFC831 CRC64;  
 Query Match 72.7%; Score 80; DB 2; Length 1477;  
 Best Local Similarity 68.2%; Pred. No. 0.001; Mismatches 4; Indels 0; Gaps 0;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VPSYSPARADSEVDLIRDI 22  
 DB 603 IPNYSFVRADSEVQTIAQII 624  
 RESULT 6  
 Q9EZHS PRELIMINARY; PRT; 1508 AA.  
 ID Q9EZHS;  
 AC Q9EZHS;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DR Dextranucrase DsrB742.  
 GN DSRB742.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF294469; AAG38021.1; -  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding; 1; 13.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;  
 Query Match 72.7%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 68.2%; Pred. No. 0.0011; Mismatches 4; Indels 0; Gaps 0;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VPSYSPARADSEVDLIRDI 22  
 DB 634 IPNYSFVRADSEVQTIAQII 655  
 RESULT 7  
 O52224 PRELIMINARY; PRT; 1508 AA.  
 AC O52224;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Glucosyltransferase (EC 2.4.1.5).  
 GN DSRB.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1299;  
 RA Monchois V., Renaud-Simeon M., Monan P., Willemot R.M.;  
 RT "Cloning and sequencing of a gene coding for an extracellular  
 RT dextranucrase (DSRB) from Leuconostoc mesenteroides NRRL B-1299  
 RT synthesizing only a (1-6) glucan."  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 DR EMBL: AF030129; AAB95453.1; -  
 DR PIR: T31098.  
 DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding; 1; 13.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;  
 Query Match 72.7%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 68.2%; Pred. No. 0.0011; Mismatches 4; Indels 0; Gaps 0;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VPSYSPARADSEVDLIRDI 22  
 DB 634 IPNYSFVRADSEVQTIAQII 655  
 RESULT 8  
 Q9LCU7 PRELIMINARY; PRT; 1016 AA.  
 ID Q9LCU7;  
 AC Q9LCU7;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE Dextranucrase.  
 GN DSRT.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512F;  
 RX MEDLINE=20169623; PubMed=10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.,  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
 RT mesenteroides NRRL B-512F."  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1;  
 DR HSSP; P06278; IVUS.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR SEQUENCE 1016 AA; 110344 MW; 8896EFD13CCB47 CRC64;  
 SQ

Query Match 71.8%; Score 79; DB 2; Length 1016;  
 Best Local Similarity 71.4%; Pred. No. 0.00099;  
 Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPYSFARAHSEVQDII 22  
 DB 625 IPNYSFVRAHSEVQVIAQIV 645

RESULT 9  
 ID Q9ZAR4 PRELIMINARY; PRT; 1527 AA.  
 AC Q9ZAR4;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Dextranucrase.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512-F;  
 RA Bhatnagar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
 RT Leuconostoc mesenteroides NRRL B-512F."  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U81374; AAD10952.1;  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 12.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;  
 SQ

Query Match 71.8%; Score 79; DB 2; Length 1527;  
 Best Local Similarity 63.6%; Pred. No. 0.0016;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPYSFARAHSEVQDII 22  
 DB 652 IPNYSFVRAHSEVQVIAQIV 673

RESULT 10  
 ID Q8KRE1 PRELIMINARY; PRT; 1527 AA.  
 AC Q8KRE1;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE Dextranucrase Dextr (EC 2.4.1.5).  
 GN DSRT.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Neubauer H., Bauche A., Mollet B.;  
 RT "Isolation and characterization of the dextranucrase Dextr of  
 RT Leuconostoc mesenteroides Lcc4."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY017384; AAG61158.1;  
 DR GO; GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 12.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase; Glycosyltransferase.  
 DR SEQUENCE 1527 AA; 169835 MW; F9D0DE220BD89668 CRC64;  
 SQ

Query Match 71.8%; Score 79; DB 2; Length 1527;  
 Best Local Similarity 63.6%; Pred. No. 0.0016;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPYSFARAHSEVQDII 22  
 DB 652 IPNYSFVRAHSEVQVIAQIV 673

RESULT 11  
 ID Q8KZL5 PRELIMINARY; PRT; 1554 AA.  
 AC Q8KZL5;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase.  
 GN GTFU.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21958684; PubMed=11960691;  
 RA Hanada N., Fukushima K., Nomura Y., Sempuku H., Hayakawa M.,  
 RA Matsuda H., Shirota T., Abiko Y.;  
 RT "Cloning and nucleotide sequence analysis of the Streptococcus  
 RT sobrinus gtfu gene that produces a highly branched water-soluble  
 RT glucan."  
 RL Biochim. Biophys. Acta 1570:75-79(2002).  
 DR EMBL; AB089438; BAC07265.1;  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase.  
 DR SEQUENCE 1554 AA; 171676 MW; 6981BC1DAE24A73 CRC64;  
 SQ

Query Match 70.9%; Score 78; DB 2; Length 1554;  
 Best Local Similarity 63.6%; Pred. No. 0.0023;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPYSFARAHSEVQDII 22  
 DB 557 IPNYSFVRAHSEVQVIAQIV 578



```
RESULT 12
Q9WKJ5 PRELIMINARY; PRT; 1512 AA.
AC Q9WKJ5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DR GTF-S.
GN GTF.
OS Streptococcus criceti.
OC Bacteri; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase (gifs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Pfamid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C8C601FC14 CRC64;

QY Query Match 67.3%; Score 74; DB 2; Length 1512;
Best Local Similarity 71.4%; Pred. No. 0.0097;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 2 PSYSEPARAHDSVQDILRDII 22
560 PSYVFRADHSEVQTVIAQII 580

RESULT 13
Q8G9Q2 PRELIMINARY; PRT; 2835 AA.
AC Q8G9Q2;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DR Dextranucrase (EC 2.4.1.5) (Fragment).
GN DEXRE.
OS Leuconostoc mesenteroides.
OC Bacteri; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=2221661; PubMed=12270834;
RA Bozomnet S., Dolis-Lafarge M., Fabre E., Pizant S., Remaud-Simeon M.,
Mondan P., Willemet R.M.;
RT "Molecular characterisation of DSR-E, an alpha-1,2 linkage
RT synthesising dextranucrase with two catalytic domains.";
RN [1]
RP Bacterioid. 184:5753-5761(2002).
DR EMBL; AJ430204; CAD22883.1; -.
DR GO; GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 20.
DR Pfam; PF02324; Glyco_hydro_70; 2.
DR Transferase; Glycosyltransferase.
FT NON TER 1
SQ SEQUENCE 2835 AA; 313264 MW; D03262CDD73539D CRC64;

QY Query Match 66.4%; Score 73; DB 2; Length 2835;
Best Local Similarity 66.7%; Pred. No. 0.028;
Matches 14; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 PSYSEPARAHDSVQDILRDII 22
Db 629 PNVAFFIRADHSEVQTVIAQII 649

RESULT 14
Q9WKJ4 PRELIMINARY; PRT; 1338 AA.
AC Q9WKJ4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DR GTF-S.
GN GTF.
OS Streptococcus criceti.
OC Bacteri; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase (gifs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77236.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Pfamid.
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

QY Query Match 63.6%; Score 70; DB 2; Length 1338;
Best Local Similarity 68.2%; Pred. No. 0.037;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 1 PSYSEPARAHDSVQDILRDII 22
509 VDNVYFIRADHSEVQTVIAQII 530

RESULT 15
Q9LCH3 PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DR Glucosyltransferase.
GN GTFR.
OS Streptococcus oralis.
OC Bacteri; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC10557;
RA Fujitara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RT encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 16.
DR Pfam; PF02324; Glyco_hydro_70; 1.
```

KW Transferrase. 1575 AA; 176792 MW; 772A26E4D7C2B543 CRC64;  
SQ SEQUENCE

Query Match 63.6%; Score 70; DB 2; Length 1575;  
Best Local Similarity 70.0%; Pred. No. 0.045;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPARADSEVQDLIRDI 22  
DB 619 NYIFVRAHDSVQTVIADI 638

RESULT 16  
Q54178 PRELIMINARY; PRT; 1577 AA.

ID 054178; 054247;  
AC 054178; 054247;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Glucosyltransferase.  
GN GTFG.

OS Streptococcus gordonii Challis.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OC NCBI\_TaxID=29390;  
CX [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=CHALLIS;  
RC MEDLINE=96157084; PubMed=8586195;  
RX Vicherman M.M., Sulavik M.C., Clewell D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
RT phase variants.";  
RT Dev. Biol. Stand. 85:309-314 (1995).  
RN [2]  
RN SEQUENCE OF 1-96 FROM N.A.

RC STRAIN=CHALLIS;  
RX MEDLINE=92276337; PubMed=1534326;  
RA Sulavik M.C., Tardif G., Clewell D.B.;  
RT "Identification of a gene, rgs, which regulates expression of  
RT glucosyltransferase and influences the Spp phenotype of Streptococcus  
RT gordonii Challis.";  
RT J. Bacteriol. 174:3577-3586 (1992).  
DR EMBL; U12643; AAC3483.1; -;  
DR EMBL; M89776; AAA26969.1; -;  
DR PIR; B41898; B41898.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CM binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CM\_binding\_1; 13.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferrase.  
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5B08D18 CRC64;

Query Match 63.6%; Score 70; DB 2; Length 1577;  
Best Local Similarity 70.0%; Pred. No. 0.045;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPARADSEVQDLIRDI 22  
DB 621 NYIFVRAHDSVQTVIADI 640

RESULT 17

ID 000599 PRELIMINARY; PRT; 1599 AA.

AC 000599;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Glucosyltransferase S precursor (EC 2.4.1.5) (GTF) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN GTFK.

OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.

OC NCBI\_TaxID=1304;  
CX [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25975;  
RX MEDLINE=93381463; PubMed=8371114;  
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;  
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and  
RT evolution of the gtf genes of oral streptococci.";  
RT J. Gen. Microbiol. 139:1511-1522 (1993).  
RN [2]  
RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 25975;  
RX MEDLINE=92148377; PubMed=1838391.  
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;  
RT "Molecular characterization of a cluster of at least two  
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
RT J. Gen. Microbiol. 137:2577-2593 (1991).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO  
CC PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF  
CC THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL) (N) = D-  
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL) (N+1).  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- SIMILARITY: TO REGIONS OF BARLEY AND BACILLUS AMYLOLIQUEFACIENS  
CC ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.

CC EMBL; Z11872; CAAT7898.1; -;  
CC EMBL; Z11873; CAAT7901.1; -;  
CC EMBL; M64111; AAA26897.1; -;  
CC PIR; S22737; S22737.  
DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CM binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CM\_binding\_1; 13.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferrase; Glucosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 42  
FT CHAIN 43 1599  
FT SEQUENCE 1599 AA; 176480 MW; 2AB7869E152B707 CRC64;

Query Match 63.6%; Score 70; DB 2; Length 1599;  
Best Local Similarity 70.0%; Pred. No. 0.045;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPARADSEVQDLIRDI 22  
DB 574 TYLFRADHSEVQTVIADI 593

RESULT 18  
068542 PRELIMINARY; PRT; 1449 AA.

ID 068542;  
AC 068542;  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Glucosyltransferase N (Fragment).  
GN GTFN.

OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OC NCBI\_TaxID=1304;  
CX [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=V1477;  
RA Jaffe R.I.;  
RT "Streptococcus salivarius V1477 gtfN".  
RT Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF049609; AAC05156.1; -.  
 DR PIR: T30552; T30552.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 7.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase.  
 DR NON\_TER  
 FT SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;  
 SQ

Query Match 61.8%; Score 68; DB 2; Length 1449;  
 Best Local Similarity 65.0%; Pred. No. 0.085;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDLIRII 22  
 :|||:|||||:|||||  
 Db 609 NYAFVRAHSEVQSIIGQII 628

RESULT 19  
 Q05264 PRELIMINARY; PRT; 1449 AA.  
 AC Q05264;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase precursor.  
 GN GTF.  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1304;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=95122197; PubMed=7822030;  
 RX Simpson C.L., Giffard P.M., Jacques N.A.;  
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
 coding for primer-independent glucosyltransferases.";  
 RT Infect. Immun. 63:609-621 (1995).  
 RL EMBL: L35495; AAC41412.1; -.  
 DR PIR: T30857; T30857.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 7.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Signal; Transferase  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.  
 FT SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;  
 SQ

Query Match 61.8%; Score 68; DB 2; Length 1449;  
 Best Local Similarity 65.0%; Pred. No. 0.085;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDLIRII 22  
 :|||:|||||:|||||  
 Db 609 NYAFVRAHSEVQSIIGQII 628

RESULT 20  
 Q09E05 PRELIMINARY; PRT; 2057 AA.  
 ID Q09E05;  
 AC Q09E05;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Alternansucrase (EC 2.4.1.140).  
 GN ASR.  
 OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OC NCBI\_TaxID=1245;  
 RN [1]  
 RE SEQUENCE FROM N.A.  
 RC STRAIN=NRL B-1355;  
 RX MEDLINE=20080809; PubMed=10612736;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,  
 RA Willemot R.M., Monsan P.;  
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
 RT glucosyltransferase from Leuconostoc mesenteroides NRL B-1355.";  
 RL FEMS Microbiol. Lett. 182:81-85 (2000).  
 DR EMBL: AJ250173; CAB65910.2; -.  
 DR GO: GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 11.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Glycosyltransferase; Transferase.  
 FT SEQUENCE 2057 AA; 228987 MW; 62BC9385D9A11BE CRC64;  
 SQ

Query Match 61.8%; Score 68; DB 2; Length 2057;  
 Best Local Similarity 63.2%; Pred. No. 0.13;  
 Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYFARAHSEVQDLIR 19  
 :|||:|||||:|||||  
 Db 757 IPNYSFVRAHSDYDADPIR 775

RESULT 21  
 Q00600 PRELIMINARY; PRT; 1518 AA.  
 ID Q00600;  
 AC Q00600;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase I (EC 2.4.1.5) (GTF) (Dextranucrase) (Sucrose 6-  
 DE glucosyltransferase).  
 GN GTF.  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1304;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 25975;  
 RX MEDLINE=92148377; PubMed=1838391;  
 RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;  
 RT "Molecular characterization of a cluster of at least two  
 RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
 RT J. Gen. Microbiol. 137:2577-2593 (1991).  
 RL U. Gen. Microbiol. 137:2577-2593 (1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO  
 CC PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF  
 CC THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL) (N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL) (N+1).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- DISEASE: DENTAL CARIES.  
 DR EMBL: Z11873; CAAY7900.1; -.  
 DR PIR: A44811; A44811.  
 DR GO: GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 11.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase; Glycosyltransferase; Repeat; Dental caries.  
 FT DOMAIN 1307 1482 6 DIRECT REPEATS.  
 FT REPEAT 1307 1338 REPEAT 1.  
 FT REPEAT 1339 1352 REPEAT 2.

FT REPEAT 1372 1403 REPEAT 3.  
FT REPEAT 1404 1417 REPEAT 4.  
FT REPEAT 1437 1468 REPEAT 5.  
FT REPEAT 1469 1482 REPEAT 6.  
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;

Query Match 59.1%; Score 65; DB 2; Length 1518;  
Best Local Similarity 60.0%; Pred. No. 0.27;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFARADSEVODLIRII 22  
Db 604 NYIFRAHSDNNVODIIAEII 623

RESULT 22  
055265 PRELIMINARY; PRT; 1577 AA.  
AC 055265  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Glucosyltransferase precursor.  
GN GTEM.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95122197; PubMed=7822030;  
RA Simpson C.L., Giffard P.M., Jacques N.A.;  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
RL infect. Immun. 63:609-621 (1995).  
DR EMBL: L35928; AAC41413.1; -.  
DR PIR: T30858; T30858.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro: IPR004829; Ceufase; antigen.  
DR InterPro: IPR002479; CW binding.  
DR InterPro: IPR00318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW binding\_1; 10.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
DR ProDom: PD153432; Ceufase\_antigen; 1.  
DR SIGNAL: 1  
FT SIGNAL 38 POTENTIAL.  
FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.  
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7B53 CRC64;

Query Match 56.4%; Score 62; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.86;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 3 SYSFARADSEVODLIRII 22  
Db 661 NYIFRAHSDSEVQAVIANII 680

RESULT 23  
0921X9 PRELIMINARY; PRT; 93 AA.  
AC 0921X9  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochetes; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
SQ STRAIN=N40;

RA Feng S., Hodzic E., Barthold S.W.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF006036; AAD01254.1; -.  
KM Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 93 AA; 10129 MW; 5EPD5DBE404AFBA8A CRC64;

Query Match 47.3%; Score 52; DB 2; Length 93;  
Best Local Similarity 58.8%; Pred. No. 1.5;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PSYFARADSEVODII 18  
Db 70 PSFSPRAHSSALLEII 86

RESULT 24  
08026 PRELIMINARY; PRT; 1554 AA.  
AC 08026  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE MYO2.  
OS Cryptococcus neoformans var. grubii.  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
OX NCBI\_TaxID=178876;  
RN [1]  
RP SEQUENCE OF 940-1554 FROM N.A.  
RX STRAIN=H99;  
RX MEDLINE=20570501; PubMed=11121047;  
RA Lengeler K.B., Wang P., Cox G.M., Perfect J.R., Heitman J.;  
RT "Identification of the MATa mating-type locus of Cryptococcus  
RL neoformans reveals a serotype A MATa strain thought to have been  
RT extinct.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:14455-14460 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H99;  
RX MEDLINE=22341086; PubMed=12455690;  
RA Lengeler K.B., Fox D.S., Fraser J.A., Allen A., Forrester K.,  
RL Dietrich F.S., Heitman J.;  
RT "Mating-type locus of Cryptococcus neoformans: a step in the evolution  
RL of sex chromosomes.";  
RL Eukaryot. Cell 1:704-718 (2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H99;  
RX Wang P., Heitman J.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H99;  
RX Dietrich F.S., Fox D.S., Fraser J.A., Allen A., Forrester K.,  
RL Lengeler K.B., Heitman J.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF542529; AAN75169.1; -.  
DR GO: GO:0016459; C:myosin; IEA.  
DR GO: GO:0005524; F:ATP binding; IEA.  
DR GO: GO:0003774; F:motor activity; IEA.  
DR InterPro: IPR002710; DIL.  
DR InterPro: IPR000048; IQ\_region.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF01843; DIL; 1.  
DR Pfam: PF00612; IQ; 6.  
DR Pfam: PF00063; myosin\_head; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000351; myosin\_head; 1.  
DR SMART: SM00015; IQ; 6.  
DR SMART: SM00242; MYSC; 1.  
SQ SEQUENCE 1554 AA; 174915 MW; ED8C245456488335 CRC64;

Query Match 45.5%; Score 50; DB 3; Length 1554;  
 Best Local Similarity 47.4%; Pred. No. 71;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPSSYPARAHSEVQDLIR 19

DB 1253 VPGYDFSNHSDSDWRCYIR 1271

RESULT 25

Q8U0W2 PRELIMINARY; PRT; 1567 AA.

AC Q8U0W2; 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

OS Cryptococcus neoformans var. neoformans.

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;

OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.

OX NCBI\_TaxID=40410;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=JEC21;

RA MEDLINE=2243086; PubMed=12455690;

RA Lengeler K.B., Fox D.S., Fraser J.A., Allen A., Forrester K.,

RA Dietrich F.S., Heitman J.;

RT "Mating-type locus of Cryptococcus neoformans: a step in the evolution

of sex chromosomes.";

RT Eukaryot. Cell 1:704-718(2002).

DR EMBL; AF542531; AAN75723.1; -.

DR GO; GO:0016459; C:myosin; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0003774; F:motor activity; IEA.

DR InterPro; IPR002710; DIL.

DR InterPro; IPR000048; IQ\_region.

DR InterPro; IPR001609; myosin\_head.

DR Pfam; PF01843; DIL; 1.

DR Pfam; PF00062; IQ; 6.

DR PRINTS; PR00193; MYOSINHEAD.

DR PRODOM; PD000355; myosin\_head; 1.

DR SMART; SM00015; IQ; 6.

DR SMART; SM00242; MYSC; 1.

SQ SEQUENCE 1567 AA; 176363 MW; D875A36BB8EBAF2 CRC64;

Query Match 45.5%; Score 50; DB 3; Length 1567;

Best Local Similarity 47.4%; Pred. No. 72;

Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPSSYPARAHSEVQDLIR 19

DB 1266 VPGYDFSNHSDSDWRCYIR 1284

RESULT 26

P74028 PRELIMINARY; PRT; 108 AA.

AC P74028; 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)

DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Hypothetical protein s11219.

GN S11219.

OS Synecocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.

OX NCBI\_TaxID=1148;

RN [1] SEQUENCE FROM N.A.

RP MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kocani H., Tanaka A., Asamizu B., Nakamura Y.,

RA Miyajima N., Hirosewa M., Sugiyama T., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shingo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synecocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";

RT DNA Res. 3:109-136(1996).

RL EMBL; D90911; BAA18101.1; -.

DR PIR; S75540; S75540.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 108 AA; 12030 MW; 589F5770326C0DBF CRC64;

Query Match 44.5%; Score 49; DB 16; Length 108;

Best Local Similarity 38.9%; Pred. No. 5.3;

Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 5 SFARAHSEVQDLIRDI 22

DB 34 NWARAHDSLDVVGQRL 51

RESULT 27

O8A338 PRELIMINARY; PRT; 303 AA.

AC O8A338; 01-JUN-2003 (TREMBLrel. 24, Created)

DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)

DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Putative oxidoreductase.

GN B73117.

OS Bacteroides thetaiotaomicron.

OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;

OC Bacteroidaceae; Bacteroides.

OX NCBI\_TaxID=818;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=VP1-5482 / ATCC 29148;

RA MEDLINE=22550858; PubMed=12663928;

RA Xu J., Bjursell M.K., Hultro J., Deng S., Carmichael L.K.,

RA Chiang H.C., Hooper L.V., Gordon J.I.;

RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";

RL EMBL; AE016939; AAO78823.1; -.

DR GO; GO:0006777; F:2-dehydropanoate 2-reductase activity; IEA.

DR GO; GO:0000036; F:acyl carrier activity; IEA.

DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.

DR GO; GO:0006206; P:pyrimidine base metabolism; IEA.

DR InterPro; IPR008927; 6DGDH\_C like.

DR InterPro; IPR003231; Acyl\_carrier.

DR InterPro; IPR003710; Apba.

DR Pfam; PF02558; Apba; 1.

DR PRODOM; PD000887; Acyl\_carrier; 1.

DR TIGRFAMs; TIGR00745; apba\_pam; 1.

KM Complete proteome.

SQ SEQUENCE 303 AA; 32667 MW; 72DSD7250673271E CRC64;

Query Match 44.5%; Score 49; DB 16; Length 303;

Best Local Similarity 56.2%; Pred. No. 17;

Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 ARAHSEVQDLIRDI 22

DB 265 ARGHESRIQGLPDM 280

RESULT 28

Q9RNP7 PRELIMINARY; PRT; 51 AA.

AC Q9RNP7; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE VCO26.

GN VCO26 OR VCA0471.

```

OS  Vibrio cholerae.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC  Vibrionaceae; Vibrio.
OX  NCBI_TaxID=666;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Classical 569B / ATCC 25870 / Serotype O1; TRANSPOSON=MINVC;
RA  Clark C.A., Manning P.A.;
RT  "The Vibrio cholera Mega-integron.";
RL  Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=El Tor N16961 / Serotype O1;
RA  Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA  Heidelberg J.F., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA  Dodson R.J., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA  Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA  Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA  McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA  Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA  Fraser C.M.;
RT  "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT  cholerae.";
RL  Nature 406:477-483 (2000).
DR  EMBL; AF179593; AAF05323.1; -.
DR  EMBL; AE004379; AAF96375.1; -.
DR  PIR; G82455; G82455.
DR  TIGR; VCA0471; -.
DR  TIGR; VCO26; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 51 AA; 5824 MW; 7580F3D7AB54E504 CRC64;

Query Match 42.7%; Score 47; DB 16; Length 51;
Best Local Similarity 50.0%; Pred. No. 4.9;
Matches 11; Conservative 6; Mismatches 3; Indels 2; Gaps 1;

QY 1 VPSYFARAHDSVQDLIRDI 22
DB 15 VPAYSAIR--NSIRIRLEKEII 34

RESULT 29
Q983M9 PRELIMINARY; PRT; 462 AA.
ID Q983M9;
AC Q983M9;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Protease.
GN M18255.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Pseudomonadaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAF303099;
RX MEDLINE=21062930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matsumoto A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338 (2000).
DR EMBL; AP003013; BAB53851.1; -.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001431; Peptidase_M16_C.
DR InterPro; IPR007863; Peptidase_M16_C.
DR Pfam; PF00675; Peptidase_M16; 1.
DR Pfam; PF05193; Peptidase_M16_C; 1.

```

```

KW Complete proteome.
SQ SEQUENCE 462 AA; 50952 MW; 1BC7B3DDFDCB28C4 CRC64;

Query Match 42.7%; Score 47; DB 16; Length 462;
Best Local Similarity 40.9%; Pred. No. 56;
Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPSYFARAHDSVQDLIRDI 22
DB 290 VPSYHTKPKPGABALDLIAEL 311

RESULT 30
Q9PR58 PRELIMINARY; PRT; 597 AA.
ID Q9PR58;
AC Q9PR58;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE DNA polymerase III gamma-tau subunits.
GN DNA OR U0087.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Seovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Caswell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum.";
RL Nature 407:757-762 (2000).
DR EMBL; AE002108; AAF30492.1; -.
DR GO; GO:0005663; C:DNA replication factor C complex; IEA.
DR GO; GO:0005624; F:DNA binding; IEA.
DR GO; GO:0005677; F:DNA binding; IEA.
DR GO; GO:000166; F:nucleotide binding; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR003593; AAA_Arpase.
DR InterPro; IPR003959; AAA_Arpase.
DR InterPro; IPR001270; Chaperin_c1pA/B.
DR InterPro; IPR000862; RfcDomain.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 597 AA; 69038 MW; C61CC497BF113D1E CRC64;

Query Match 42.7%; Score 47; DB 16; Length 597;
Best Local Similarity 47.4%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 VPSFARAHDSVQDLIRDI 22
DB 173 YDFKRLNSELQELIDIL 191

```

Search completed: May 4, 2004, 09:12:32  
Job time : 36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:07:31 ; Search time 14 Seconds  
(without alignments)  
81.127 Million cell updates/sec

Title: US-09-290-049a-17  
Perfect score: 110  
Sequence: 1 VPSYSPARAHSEVQDLIRDT 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Issued Patents AA:\*  
1: /cgm2\_6/prodata/2/iaa/5A\_COMB.pep.\*  
2: /cgm2\_6/prodata/2/iaa/5B\_COMB.pep.\*  
3: /cgm2\_6/prodata/2/iaa/6A\_COMB.pep.\*  
4: /cgm2\_6/prodata/2/iaa/6B\_COMB.pep.\*  
5: /cgm2\_6/prodata/2/iaa/PCUTS\_COMB.pep.\*  
6: /cgm2\_6/prodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	90.9	1375	3 US-09-210-361-4	Sequence 4, Appli
2	100	90.9	1375	4 US-09-740-374-4	Sequence 4, Appli
3	99	90.0	1475	3 US-09-007-999-2	Sequence 2, Appli
4	99	90.0	1475	3 US-09-210-361-2	Sequence 2, Appli
5	99	90.0	1475	4 US-09-740-374-2	Sequence 2, Appli
6	79	71.8	523	4 US-09-604-957-5	Sequence 5, Appli
7	68	61.8	535	4 US-09-604-957-7	Sequence 7, Appli
8	68	61.8	1278	4 US-09-604-957-3	Sequence 3, Appli
9	68	61.8	2057	4 US-09-499-203-2	Sequence 6, Appli
10	63	57.3	584	4 US-09-604-957-6	Sequence 6, Appli
11	62	56.4	545	4 US-09-604-957-4	Sequence 4, Appli
12	62	56.4	1430	3 US-09-008-172-2	Sequence 2, Appli
13	62	56.4	1430	3 US-09-210-361-6	Sequence 6, Appli
14	62	56.4	1430	4 US-09-740-274-6	Sequence 6, Appli
15	62	56.4	1577	2 US-08-793-824-2	Sequence 2, Appli
16	45	40.9	2627	2 US-08-751-189-3	Sequence 3, Appli
17	45	40.9	2627	2 US-09-060-836-3	Sequence 3, Appli
18	45	40.9	2627	2 US-09-184-445-3	Sequence 3, Appli
19	44.5	40.5	501	4 US-09-134-001C-4115	Sequence 4115, Ap
20	44	40.0	309	4 US-09-345-473B-37	Sequence 37, Appli
21	43.5	39.5	484	3 US-08-913-578-2	Sequence 2, Appli
22	43.5	39.5	484	3 US-08-785-427-2	Sequence 2, Appli
23	43	39.1	639	2 US-08-557-309B-37	Sequence 37, Appli
24	43	39.1	639	3 US-08-834-306-37	Sequence 37, Appli
25	43	39.1	639	3 US-08-993-674A-37	Sequence 37, Appli
26	43	39.1	639	4 US-09-256-976-37	Sequence 37, Appli
27	42	38.2	196	2 US-08-684-024-2	Sequence 2, Appli

28	42	38.2	196	2 US-08-684-024-9	Sequence 9, Appli
29	42	38.2	196	3 US-09-145-868-2	Sequence 2, Appli
30	42	38.2	196	3 US-09-145-868-9	Sequence 9, Appli
31	42	38.2	565	4 US-09-107-532A-4217	Sequence 4217, Ap
32	41	37.3	639	4 US-09-509-814A-4	Sequence 4, Appli
33	41	37.3	933	1 US-08-370-193A-8	Sequence 8, Appli
34	41	37.3	933	4 US-09-271-438A-9	Sequence 9, Appli
35	41	37.3	933	4 US-10-078-107-5	Sequence 5, Appli
36	41	37.3	933	4 US-10-077-751-5	Sequence 5, Appli
37	41	37.3	933	4 US-09-271-438A-3	Sequence 3, Appli
38	41	37.3	933	4 US-09-271-438A-8	Sequence 8, Appli
39	41	37.3	933	4 US-10-078-107-1	Sequence 1, Appli
40	41	37.3	933	4 US-10-077-751-1	Sequence 1, Appli
41	41	37.3	938	4 US-09-489-039A-13504	Sequence 13504, A
42	40	36.4	43	3 US-08-679-006-2	Sequence 2, Appli
43	40	36.4	75	4 US-09-107-532A-7020	Sequence 7020, Ap
44	40	36.4	191	4 US-09-511-024A-9	Sequence 9, Appli
45	40	36.4	340	4 US-09-328-352-7834	Sequence 7834, Ap
46	40	36.4	392	4 US-09-424-978B-39	Sequence 39, Appli
47	40	36.4	1068	3 US-08-390-874C-11	Sequence 11, Appli
48	40	36.4	1068	4 US-09-265-772-11	Sequence 11, Appli
49	40	36.4	1069	2 US-08-162-081B-37	Sequence 37, Appli
50	40	36.4	1069	2 US-08-780-872-37	Sequence 37, Appli
51	40	36.4	1069	3 US-09-085-957-37	Sequence 37, Appli
52	40	36.4	1080	2 US-08-162-081B-36	Sequence 36, Appli
53	40	36.4	1080	2 US-08-780-872-36	Sequence 36, Appli
54	40	36.4	1080	3 US-09-085-957-36	Sequence 36, Appli
55	40	36.4	1454	3 US-08-392-459-22	Sequence 22, Appli
56	40	36.4	1454	4 US-08-392-459-26	Sequence 26, Appli
57	40	36.4	1454	4 US-09-854-799-22	Sequence 22, Appli
58	40	36.4	1454	5 PCT-US91-08525-22	Sequence 22, Appli
59	40	36.4	1454	5 PCT-US91-08525-26	Sequence 26, Appli
60	40	36.4	1454	5 PCT-US93-04384-2	Sequence 2, Appli
61	40	36.4	1454	5 PCT-US93-04384-8	Sequence 8, Appli
62	40	36.4	1454	5 PCT-US93-04384-16	Sequence 16, Appli
63	40	36.4	1454	5 PCT-US93-04384-43	Sequence 43, Appli
64	40	36.4	1454	5 PCT-US93-04384-45	Sequence 45, Appli
65	40	36.4	1454	5 PCT-US93-04384-48	Sequence 48, Appli
66	40	36.4	1454	5 PCT-US93-04384-7	Sequence 7, Appli
67	40	36.4	1497	1 US-08-623-679-7	Sequence 7, Appli
68	40	36.4	1497	3 US-08-933-030-7	Sequence 7, Appli
69	40	36.4	1497	3 US-09-181-070-7	Sequence 7, Appli
70	40	36.4	1497	4 US-09-534-242-7	Sequence 7, Appli
71	40	36.4	1497	4 US-09-454-854-7	Sequence 7, Appli
72	40	36.4	1497	4 US-09-164-671-7	Sequence 7, Appli
73	40	36.4	1533	1 US-08-623-679-9	Sequence 9, Appli
74	40	36.4	1533	3 US-08-933-774-9	Sequence 9, Appli
75	40	36.4	1533	3 US-09-181-030-9	Sequence 9, Appli
76	40	36.4	1533	4 US-09-534-242-9	Sequence 9, Appli
77	40	36.4	1533	4 US-09-454-854-9	Sequence 9, Appli
78	40	36.4	1533	4 US-09-164-671-9	Sequence 9, Appli
79	39.5	35.9	387	4 US-09-252-931A-31249	Sequence 31249, A
80	39.5	35.9	387	4 US-09-638-937-9	Sequence 9, Appli
81	39.5	35.5	126	3 US-08-331-625A-9	Sequence 9, Appli
82	39	35.5	126	4 US-09-494-151-9	Sequence 9, Appli
83	39	35.5	126	4 US-09-972-484-9	Sequence 9, Appli
84	39	35.5	184	1 US-08-353-550-2	Sequence 2, Appli
85	39	35.5	184	1 US-08-551-687-2	Sequence 2, Appli
86	39	35.5	251	2 US-08-331-625A-59	Sequence 59, Appli
87	39	35.5	251	4 US-09-494-151-59	Sequence 59, Appli
88	39	35.5	251	4 US-09-972-484-59	Sequence 59, Appli
89	39	35.5	282	4 US-09-172-952-20	Sequence 20, Appli
90	39	35.5	364	4 US-09-489-039A-8909	Sequence 8909, Ap
91	39	35.5	428	1 US-08-353-550-1	Sequence 1, Appli
92	39	35.5	428	2 US-08-551-687-1	Sequence 1, Appli
93	39	35.5	431	3 US-08-807-342B-5	Sequence 5, Appli
94	39	35.5	441	4 US-09-540-236-2983	Sequence 2883, Ap
95	39	35.5	469	1 US-08-353-550-6	Sequence 6, Appli
96	39	35.5	469	2 US-08-551-687-6	Sequence 6, Appli
97	39	35.5	551	4 US-09-252-931A-25323	Sequence 25323, A
98	39	35.5	1101	3 US-08-331-625A-52	Sequence 52, Appli
99	39	35.5	1101	3 US-08-331-625A-54	Sequence 54, Appli
100	39	35.5	1101	4 US-09-494-151-52	Sequence 52, Appli



## ALIGNMENTS

RESULT 1  
US-09-210-361-4

Sequence 4, Application US/09210361  
Patent No. 6284479

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Substitutes for Modified Starches and

TITLE OF INVENTION: Latexes in Paper Manufacture

FILE REFERENCE: 0357CR

CURRENT FILING DATE: US/09/210,361

EARLIER FILING DATE: 1998-12-11

EARLIER APPLICATION NUMBER: 09/007,999

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/009,620

EARLIER FILING DATE: 1998-01-20

EARLIER APPLICATION NUMBER: 08/485,243

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/008,172

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/482,711

EARLIER FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 1375

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-210-361-4

Query Match

Best Local Similarity 90.9%; Score 100; DB 3; Length 1375;

Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 578 VPSYSFIRAHSEVQDLIRNII 599

QY 1 VPSYSFARAHSEVQDLIRDII 22

US-09-210-274-4

Sequence 4, Application US/09740274

Patent No. 6465203

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

FILE REFERENCE: 0357CRD

CURRENT FILING DATE: US/09/740,274

EARLIER FILING DATE: 2000-12-19

EARLIER APPLICATION NUMBER: 09/210,361

EARLIER FILING DATE: 1998-12-11

EARLIER APPLICATION NUMBER: 09/007,999

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/009,620

EARLIER FILING DATE: 1998-01-20

EARLIER APPLICATION NUMBER: 08/485,243

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/008,172

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/482,711

EARLIER FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-4  
Query Match  
Best Local Similarity 90.9%; Score 100; DB 4; Length 1375;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLIRDII 22  
Db 578 VPSYSFIRAHSEVQDLIRNII 599

RESULT 3

US-09-007-999-2

Sequence 2, Application US/09007999

Patent No. 6087559

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Substitutes for Modified Starch and

TITLE OF INVENTION: Latexes in Paper Manufacture

FILE REFERENCE: 0356D

CURRENT FILING DATE: US/09/007,999

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/478,704

EARLIER FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 1475

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-007-999-2

Query Match

Best Local Similarity 90.0%; Score 99; DB 3; Length 1475;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 553 VPSYSFIRAHSEVQDLIRDII 573

QY 1 VPSYSFARAHSEVQDLIRDII 22

US-09-210-361-2

Sequence 2, Application US/09210361

Patent No. 6284479

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Substitutes for Modified Starches and

TITLE OF INVENTION: Latexes in Paper Manufacture

FILE REFERENCE: 0357CR

CURRENT FILING DATE: US/09/210,361

EARLIER FILING DATE: 1998-12-11

EARLIER APPLICATION NUMBER: 09/007,999

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/009,620

EARLIER FILING DATE: 1998-01-20

EARLIER APPLICATION NUMBER: 08/485,243

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/008,172

EARLIER FILING DATE: 1998-01-16

EARLIER APPLICATION NUMBER: 08/482,711

EARLIER FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 1475

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-210-361-2

Query Match 90.0%; Score 99; DB 3; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 5,4e-08;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSYSPARADSEVODLIRDI 22  
DB 552 VPSYSPARADSEVODLIRDI 573

RESULT 5  
US-09-740-274-2  
Sequence 2, Application US/09740274  
Patent No. 6465203

GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR FILING DATE: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 90.0%; Score 99; DB 4; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 5,4e-08;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSYSPARADSEVODLIRDI 22  
DB 552 VPSYSPARADSEVODLIRDI 573

RESULT 6  
US-09-604-957-5  
Sequence 5, Application US/09604957  
Patent No. 6486314  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 523  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 71.8%; Score 79; DB 4; Length 523;  
Best Local Similarity 63.6%; Pred. No. 3,7e-05;  
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARADSEVODLIRDI 22  
DB 146 IPNYSFVARADSEVQTVIAQIV 167

RESULT 7  
US-09-604-957-7  
Sequence 7, Application US/09604957  
Patent No. 6486314

GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 535  
TYPE: PRT  
ORGANISM: Lactobacillus reuteri  
US-09-604-957-7

Query Match 61.8%; Score 68; DB 4; Length 535;  
Best Local Similarity 54.5%; Pred. No. 0.0027;  
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSPARADSEVODLIRDI 22  
DB 144 IPNYSFVARADSEVQTVIAQIV 165

RESULT 8  
US-09-604-957-3  
Sequence 3, Application US/09604957  
Patent No. 6486314  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1278  
TYPE: PRT  
ORGANISM: Lactobacillus reuteri  
US-09-604-957-3

Query Match 61.8%; Score 68; DB 4; Length 1278;  
Best Local Similarity 54.5%; Pred. No. 0.0077;  
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSPARADSEVODLIRDI 22  
DB 620 IPNYSFVARADSEVQTVIAQIV 641

```

RESULT 9
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499, 203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

```

```
QY      1 VPSYSFARAH DSEVQDLIR 19
          :|:| | | | | | : | | |
Db      757 IPNYSFVARAH DYDAQDPIR 775
```

```

RESULT 10
US-09-604-957-6
/ Sequence 6, Application US/09604957
/ Patent No. 6486314
/ GENERAL INFORMATION:
/ APPLICANT: VAN GEEL-SCHOTTEN, GERRITDINA HENDRIKA
/ APPLICANT: DIKHUIZEN, LUBBERT
/ APPLICANT: RAHROUFI, HAKIM
/ APPLICANT: LEER, ROBERT-JAN
/ TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
/ FILE REFERENCE: BO 43586
/ CURRENT APPLICATION NUMBER: US/09/604,957
/ CURRENT FILING DATE: 2000-06-28
/ PRIOR APPLICATION NUMBER: 00201871.1
/ PRIOR FILING DATE: 2000-05-25
/ NUMBER OF SEQ ID NOS: 17
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 6
/ LENGTH: 584
/ TYPE: PRT
/ ORGANISM: Leuconostoc mesenteroides
US-09-604-957-6

```

```
QY      1 VPSYSPFAHDSVQDLI 18
        :|:|:|:|:|:|:|:|
Db      167 IPNYSFVRAHDYDAQDPI 184
```

RESULT 11  
US-09-604-957-4  
Sequence 4, Application US/09604957  
Patent No. 6486314  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITJINA HENDRIKA  
APPLICANT: DIGHUIZEN, LUBBERT  
APPLICANT: RAHNOU, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

```

1 FILE REFERENCE: BO 43388
2 CURRENT APPLICATION NUMBER: US/09/604,957
3 CURRENT FILING DATE: 2000-06-28
4 PRIOR APPLICATION NUMBER: 000201871.1
5 PRIOR FILING DATE: 2000-05-25
6 NUMBER OF SEQ ID NOS: 17
7 SOFTWARE: PatentIn Ver. 2.1
8 SEQ ID NO 4
9 LENGTH: 545
10 TYPE: PR1
11 ORGANISM: Streptococcus mitis
12 US-09-604-957-4

```

```
QY      3 SYSPARAHDSVQDLIRDTI 22
        :| | | | | | | | :| |
Db      156 NYIFIRAHDSVQTVIAKII 175
```

```

RESULT 12
US-09-008-172-2
Sequence 2, Application US/09008172
Patent No. 6127602
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Strach and
FILE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 0358D
CURRENT APPLICATION NUMBER: US/09/008,172
CURRENT FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FaastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1430
TYPE: FRT
ORGANISM: Streptococcus mutans
US-09-008-172-2

```

```
QY      3 SYSFARAH DSEVQD LIRDI 22
          :| | | | | | | :| | |
Db      576 NYIFIRAH DSEVQTVIAKII 595
```

RESULT 13  
US-09-210-361-6  
? Sequence 6, Application US/09210361  
? Patent No. 6284479  
? GENERAL INFORMATION:  
? APPLICANT: Nicholas, Scott E.  
? TITLE OF INVENTION: Substitutes for Modified Starches and  
? TITLE OF INVENTION: Latexes in Paper Manufacture  
? PIR REFERENCE: 0357CR  
? CURRENT APPLICATION NUMBER: US/09/210,361  
? CURRENT FILING DATE: 1998-12-11  
? EARLIER APPLICATION NUMBER: 09/007,999  
? EARLIER FILING DATE: 1998-01-16  
? EARLIER APPLICATION NUMBER: 08/478,704  
? EARLIER FILING DATE: 1995-06-07  
? EARLIER APPLICATION NUMBER: 09/009,620  
? EARLIER FILING DATE: 1998-01-20  
? EARLIER APPLICATION NUMBER: 08/485,243  
? EARLIER FILING DATE: 1995-06-07  
? EARLIER APPLICATION NUMBER: 09/008,172  
? EARLIER FILING DATE: 1998-01-16

```

; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match      56.4%; Score 62; DB 3; Length 1430;
Best Local Similarity 65.0%; Pred. No. 0.09;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      3 SYSFARHSEVQDLIRDTI 22
DB      576 NYIFRHHSEVQTVIAXII 595

RESULT 14
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 035/CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match      56.4%; Score 62; DB 4; Length 1430;
Best Local Similarity 65.0%; Pred. No. 0.09;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      3 SYSFARHSEVQDLIRDTI 22
DB      576 NYIFRHHSEVQTVIAXII 595

RESULT 15
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
```

```

; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
US-08-793-824-2

Query Match      56.4%; Score 62; DB 2; Length 1577;
Best Local Similarity 60.0%; Pred. No. 0.1;
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      3 SYSFARHSEVQDLIRDTI 22
DB      661 NYIFRHHSEVQTVIAXII 680

RESULT 16
US-08-751-189-3
; Sequence 3, Application US/08751189
; Patent No. 5919656
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,189
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 2627 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-751-189-3

Query Match 40.9%; Score 45; DB 2; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 1.4e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDL 17  
:||||: :|||:  
DB 205 MPYSLSLGEEREVEDL 221

RESULT 17  
US-09-060-836-3  
Sequence 3, Application US/09060836

Patent No. 5981707  
GENERAL INFORMATION:  
APPLICANT: Harrington, Lea A.  
ATTORNEY/AGENT INFORMATION:  
TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Amgen, Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/060,836  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/751,189  
FILING DATE: 15-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Oleski, Nancy A.  
REGISTRATION NUMBER: 34,688  
REFERENCE/DOCKET NUMBER: A-433  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2627 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-060-836-3

Query Match 40.9%; Score 45; DB 2; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 1.4e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDL 17  
:||||: :|||:  
DB 205 MPYSLSLGEEREVEDL 221

RESULT 18  
US-09-184-445-3  
Sequence 3, Application US/09184445  
Patent No. 6174703  
GENERAL INFORMATION:  
APPLICANT: Harrington, Lea A.

APPLICANT: Robinson, Murray O.  
TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Amgen, Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/184,445  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/751,189  
FILING DATE: 15-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Oleski, Nancy A.  
REGISTRATION NUMBER: 34,688  
REFERENCE/DOCKET NUMBER: A-433  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2627 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-184-445-3

Query Match 40.9%; Score 45; DB 3; Length 2627;  
Best Local Similarity 45.5%; Pred. No. 1.4e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDL 17  
:||||: :|||:  
DB 205 MPYSLSLGEEREVEDL 221

RESULT 19  
US-09-134-001C-4115  
Sequence 4115, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4115  
LENGTH: 501  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4115

Query Match 40.5%; Score 44.5; DB 4; Length 501;  
Best Local Similarity 45.5%; Pred. No. 23;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSFARA--HDSVQDLIR 19  
||:|:|:| :||:|:|

Db 210 VPTYNFAVAIDHMQISDVIR 231

RESULT 20

US-09-345-473E-37  
Sequence 37, Application US/09345473E

Patent No. 6558903  
GENERAL INFORMATION:

APPLICANT: Hodge, Martin

TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof

FILE REFERENCE: 35800/183781

CURRENT APPLICATION NUMBER: US/09/345,473E

CURRENT FILING DATE: 1999-06-30

NUMBER OF SEQ ID NOS: 62

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 37

LENGTH: 309

TYPE: PRT

ORGANISM: Homo sapiens

US-09-345-473E-37

Query Match

Best Local Similarity 40.0%; Score 44; DB 4; Length 309;  
Best Local Similarity 38.9%; Pred. No. 16;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 5 SPARHSEVQDLIRDI 22

Db 190 SPEKADPRKEITIGCI 207

RESULT 21

US-08-913-578-2

Sequence 2, Application US/08913578

Patent No. 6218159

GENERAL INFORMATION:

APPLICANT: Hodgson, John

TITLE OF INVENTION: No. 6218159el tRNA synthetase

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/913,578

FILING DATE: 17-JAN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9601069.9

FILING DATE: 19-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31352

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 484 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-913-578-2

Query Match

Best Local Similarity 39.5%; Score 43.5; DB 3; Length 484;  
Best Local Similarity 40.9%; Pred. No. 32;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSPARA--HDSVQDLIR 19

Db 193 IPTYNFAVAIDHMQISDVIR 214

RESULT 22

US-08-785-427-2

Sequence 2, Application US/08785427

Patent No. 6238900

GENERAL INFORMATION:

APPLICANT: Hodgson, John

TITLE OF INVENTION: No. 6238900el tRNA synthetase

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/785,427

FILING DATE: 17-JAN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9601069.9

FILING DATE: 19-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31352

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 484 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-785-427-2

Query Match

Best Local Similarity 39.5%; Score 43.5; DB 3; Length 484;  
Best Local Similarity 40.9%; Pred. No. 32;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSPARA--HDSVQDLIR 19

Db 193 IPTYNFAVAIDHMQISDVIR 214

RESULT 23

US-08-557-309B-37

Sequence 37, Application US/08557309B

Patent No. 5916572

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Lodes, Michael J.

```

1 REFERENCE/DOCKET NUMBER: 210121.422C1
2 TELECOMMUNICATION INFORMATION:
3 TELEPHONE: (206) 622-4900
4 TELEFAX: (206) 682-6031
5 INFORMATION FOR SEQ ID NO: 37:
6 SEQUENCE CHARACTERISTICS:
7 LENGTH: 639 amino acids
8 TYPE: amino acid
9 STRANDEDNESS:
10 TOPOLOGY: linear
11
12 US-08-834-306-37
13
14 Query Match 39.1%; Score 43; DB 3; Length 639;
15 Best Local Similarity 52.9%; Pred. No. 55;
16 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
17
18 QY 1 VPSTSPARAHSEVODL 17
19 ||:|:|:|:|
20
21 Db 479 VPGMSKALHDAEPQL 495
22
23 RESULT 25
24 US-08-993-674A-37
25 Sequence 37, Application US/08993674A
26 Patent No. 6228372
27 GENERAL INFORMATION:
28 APPLICANT: Reed, Steven G.
29 APPLICANT: Skeily, Yasir A.W.
30 APPLICANT: Lodes, Michael J.
31 APPLICANT: Houghton, Raymond L.
32 APPLICANT: Smith, John M.
33 APPLICANT: McNeill, Patricia D.
34 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
35 PREVENTION OF T
36 NUMBER OF SEQUENCES: 81

```

```

Query Match          39.1%; Score 43; DB 3; Length 639;
Best Local Similarity 52.9%; Pred. No. 55;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY      1 VPSTSPARAHDSFVODL 17
      ||:|:|:|:|:|
Db      479 VPGMSREALHDAEFQOL 495

RESULT 25
US-08-993-674A-37
/ Sequence 37, Application US/08993674A
/ Patent No. 6228372
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Houghton, Raymond L.
/ APPLICANT: Smith, John M.
/ APPLICANT: McNeill, Patricia D.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
/ NUMBER OF SEQUENCES: 81
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED AND BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/993,674A
/ FILING DATE: 18-DEC-1997
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.422C2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031 37:
/ INFORMATION FOR SEQ ID NO: 37:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 639 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
/ US-08-993-674A-37

Query Match          39.1%; Score 43; DB 3; Length 639;
Best Local Similarity 52.9%; Pred. No. 55;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY      1 VPSTSPARAHDSFVODL 17
      ||:|:|:|:|:|
Db      479 VPGMSREALHDAEFQOL 495

```

```

RESULT 26
US-09-256-976-37
Sequence 37, Application US/09256976
Patent No. 6419933
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
FILE REFERENCE: 210121.422C3
CURRENT APPLICATION NUMBER: US/09/256,976
CURRENT FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 37
LENGTH: 639
TYPE: PRT
ORGANISM: Trypanosoma cruzi
FEATURE:
OTHER INFORMATION: where any Xaa is an independently selected amino
US-09-256-976-37
Query Match 39.1%; Score 43; DB 4; Length 639;
Best Local Similarity 52.9%; Pred. No. 55;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0.

QY 1 VPSYSPARAHSEVQDL 17
DB 479 VPGNSEALLHDAKFOOL 495

RESULT 27
US-08-684-024-2
Sequence 2, Application US/08684024
Patent No. 5834298
GENERAL INFORMATION:
APPLICANT: Beneza, Robert
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/684,024
FILING DATE: 19-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1747/46621-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 196 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

```

      ;      TOPOLOGY: linear
      ;      MOLECULE TYPE: protein
US-08-684-024-2
      Query Match          38.2%;      Score 42;      DB 2;      Length 196;
      Best Local Similarity 46.7%;      Pred. No. 20;
      Matches      7;      Conservative      4;      Mismatches      4;      Indels      0;      Gaps      0;

QY      8 RAHDSVQDLIRDI 22
      : ||::|||:|:
      Db      49 KTHDELKDYIRKIL 63

      RESULT 28
      US-08-684-024-9
      ; Sequence 9, Application US/08684024
      ; Patent No. 5834298
      ; GENERAL INFORMATION:
      ; APPLICANT: Benzeza, Robert
      ; TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2
      ; NUMBER OF SEQUENCES: 9
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSEE: Cooper & Dunham LLP
      ; STREET: 1185 Avenue of the Americas
      ; CITY: New York
      ; STATE: New York
      ; COUNTRY: U.S.A.
      ; ZIP: 10036
      ; COMPUTER READABLE FORM:
      ; MEDIUM TYPE: Floppy disk
      ; COMPUTER: IBM PC compatible
      ; OPERATING SYSTEM: PC-DOS/MS-DOS
      ; SOFTWARE: PatentIn Release #1.0, Version #1.30
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: US/08/684,024
      ; FILING DATE: 19-JUL-1996
      ; CLASSIFICATION: 435
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: White, John P.
      ; REGISTRATION NUMBER: 28,678
      ; REFERENCE/DOCKET NUMBER: 1747/46621-A
      ; TELECOMMUNICATION INFORMATION:
      ; TELEPHONE: (212) 278-0400
      ; TELEFAX: (212) 391-0526
      ; INFORMATION FOR SEQ ID NO: 9:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 196 amino acids
      ; TYPE: amino acid
      ; STRANDEDNESS: single
      ; TOPOLOGY: linear
      ; MOLECULE TYPE: protein
      ; US-08-684-024-9

      Query Match          38.2%;      Score 42;      DB 2;      Length 196;
      Best Local Similarity 46.7%;      Pred. No. 20;
      Matches      7;      Conservative      4;      Mismatches      4;      Indels      0;      Gaps      0

      RESULT 29
      US-09-145-868-2
      ; Sequence 2, Application US/09145868
      ; Patent No. 6096522
      ; GENERAL INFORMATION:
      ; APPLICANT: Benzeza, Robert
      ; TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2
      ; NUMBER OF SEQUENCES: 9
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSEE: Cooper & Dunham LLP
      ; STREET: 1185 Avenue of the Americas

```



Thu May 6 16:45:43 2004

us-09-290-049a-17.rai

Page 10

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-2

Query Match 38.2%; Score 42; DB 3; Length 196;  
Best Local Similarity 46.7%; Pred. No. 20;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 RAHSEVODLRDII 22  
DB 49 KTHDDRLDYIRKIL 63

RESULT 30  
US-09-145-868-9  
Sequence 9, Application US/09145868  
Patent No. 6096522  
GENERAL INFORMATION:  
APPLICANT: Benitez, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-9

Query Match 38.2%; Score 42; DB 3; Length 196;  
Best Local Similarity 46.7%; Pred. No. 20;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 8 RAHSEVODLRDII 22  
DB 49 KTHDDRLDYIRKIL 63

Search completed: May 4, 2004, 09:14:11  
Job time : 15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 4, 2004, 09:12:41 ; Search time 35.3333 Seconds  
(without alignments)  
172.590 Million cell updates/sec

Title: US-09-290-049a-17  
Perfect score: 110  
Sequence: 1 VPEYSFARAHSEVDLINDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	90.9	1375	9	US-09-740-274-4
2	99	90.0	1475	9	US-09-740-274-2
3	99	91.8	522	9	US-09-995-749A-11
4	68	61.8	535	9	US-09-995-749A-13
5	68	61.8	584	9	US-09-995-749A-12
6	68	61.8	1781	9	US-09-995-749A-2
7	68	61.8	2057	15	US-10-417-280A-2
8	62	56.4	545	9	US-09-995-749A-10
9	62	56.4	1430	9	US-09-740-274-6
10	47	42.7	117	12	US-10-424-599-168435
11	47	42.7	462	15	US-10-369-493-12332
12	47	42.7	597	12	US-10-282-122A-76700
13	46.5	42.3	465	12	US-10-282-122A-71700
14	46	41.8	54	12	US-10-424-599-232822
15	46	41.8	111	12	US-10-424-599-229462

16	46	41.8	170	12	US-10-424-599-159618	Sequence 159618,
17	45.5	41.4	481	9	US-09-815-242-558A	Sequence 558A, Ap
18	45.5	41.4	487	9	US-09-815-242-12456	Sequence 12456, A
19	45	40.9	166	12	US-10-424-599-264904	Sequence 264904,
20	45	40.9	171	12	US-10-424-599-184218	Sequence 184218,
21	45	40.9	362	12	US-10-424-599-158664	Sequence 158664,
22	45	40.9	365	12	US-10-424-599-250160	Sequence 250160,
23	45	40.9	377	12	US-10-424-599-153178	Sequence 153178, A
24	45	40.9	414	12	US-10-425-114-50726	Sequence 50726, A
25	45	40.9	423	12	US-10-425-114-65612	Sequence 65612, A
26	45	40.9	963	12	US-10-424-599-194106	Sequence 194106,
27	45	40.9	2630	12	US-10-334-143-41	Sequence 41, Appl
28	44.5	40.5	484	15	US-10-282-122A-71074	Sequence 71074, A
29	44	40.0	115	12	US-10-424-599-267457	Sequence 267457,
30	44	40.0	309	9	US-09-862-027-37	Sequence 37, Appl
31	44	40.0	400	12	US-10-276-774-2103	Sequence 2103, Ap
32	44	40.0	521	15	US-10-131-410-117	Sequence 117, App
33	44	40.0	564	15	US-10-131-410-180	Sequence 180, App
34	44	40.0	779	14	US-10-353-929-49	Sequence 49, Appl
35	44	40.0	1345	16	US-10-433-794-17	Sequence 17, Appl
36	43.5	39.5	484	12	US-10-282-122A-43803	Sequence 43803, A
37	43	39.1	99	12	US-10-425-114-49790	Sequence 49790, A
38	43	39.1	166	9	US-09-864-761-45372	Sequence 45372, A
39	43	39.1	174	12	US-10-424-599-139858	Sequence 139858,
40	43	39.1	197	12	US-10-424-599-270841	Sequence 270841,
41	43	39.1	255	9	US-09-895-828-454	Sequence 454, App
42	43	39.1	255	14	US-10-114-666-454	Sequence 454, App
43	43	39.1	262	9	US-09-895-828-457	Sequence 457, App
44	43	39.1	262	14	US-10-114-666-457	Sequence 457, App
45	43	39.1	279	12	US-10-424-599-211056	Sequence 211056,
46	43	39.1	479	15	US-10-369-493-20600	Sequence 20600, A
47	43	39.1	637	15	US-10-369-493-21656	Sequence 21656, A
48	43	39.1	704	14	US-10-213-990-21	Sequence 21, Appl
49	43	39.1	1025	14	US-10-055-475-2	Sequence 2, Appl
50	43	39.1	1025	14	US-10-055-475-7	Sequence 7, Appl
51	43	39.1	1025	14	US-10-055-475-9	Sequence 9, Appl
52	43	39.1	1025	14	US-10-055-475-12	Sequence 12, Appl
53	42.5	38.6	486	15	US-10-012-697-1539	Sequence 1539, Ap
54	42.5	38.6	486	9	US-09-815-242-13455	Sequence 13455, A
55	42.5	38.6	486	12	US-10-382-122A-74199	Sequence 74199, A
56	42.5	38.6	540	15	US-10-369-493-1936	Sequence 1936, Ap
57	42.5	38.2	116	15	US-10-333-006-14	Sequence 14, Appl
58	42	38.2	146	12	US-10-425-114-44862	Sequence 44862, A
59	42	38.2	170	12	US-10-425-114-61839	Sequence 61839, A
60	42	38.2	196	15	US-10-369-493-22169	Sequence 22169, A
61	42	38.2	227	9	US-09-993-108-6	Sequence 6, Appl
62	42	38.2	227	11	US-09-993-808B-6	Sequence 6, Appl
63	42	38.2	231	12	US-10-282-122A-58280	Sequence 58280, A
64	42	38.2	278	12	US-10-425-114-71211	Sequence 71211, A
65	42	38.2	377	12	US-10-424-599-194041	Sequence 194041,
66	42	38.2	547	12	US-10-425-114-65616	Sequence 65616, A
67	42	38.2	1383	14	US-10-021-955-82	Sequence 82, Appl
68	42	38.2	1383	14	US-10-021-955-86	Sequence 86, Appl
69	42	38.2	1383	14	US-10-021-955-87	Sequence 87, Appl
70	42	38.2	1389	14	US-10-021-955-79	Sequence 79, Appl
71	42	38.2	1835	15	US-10-369-493-1489	Sequence 1489, Ap
72	42	38.2	362	15	US-10-369-493-8187	Sequence 8187, App
73	41.5	37.7	482	12	US-10-267-502-218	Sequence 218, App
74	41	37.3	66	12	US-10-424-599-252560	Sequence 252560,
75	41	37.3	79	12	US-10-424-599-284550	Sequence 284550,
76	41	37.3	119	12	US-10-424-599-172541	Sequence 172541,
77	41	37.3	123	12	US-10-424-599-150164	Sequence 150164,
78	41	37.3	133	12	US-10-424-599-218824	Sequence 218824,
79	41	37.3	139	12	US-10-424-599-218824	Sequence 218824,
80	41	37.3	152	12	US-10-424-599-266908	Sequence 266908,
81	41	37.3	172	12	US-10-424-599-250166	Sequence 250166,
82	41	37.3	202	12	US-10-424-599-250163	Sequence 250163,
83	41	37.3	205	12	US-10-424-599-250163	Sequence 250163,
84	41	37.3	211	12	US-10-425-114-77166	Sequence 77166, A
85	41	37.3	251	12	US-10-424-599-279743	Sequence 279743, A
86	41	37.3	265	12	US-10-282-122A-51087	Sequence 51087, A
87	41	37.3	303	12	US-10-282-122A-52954	Sequence 52954, A
88	41	37.3	364	12	US-10-425-114-52806	Sequence 52806, A
					US-10-424-599-266907	Sequence 266907,

89 41 37.3 364 12 US-10-425-114-66209 Sequence 66209, A  
90 41 37.3 371 12 US-10-425-114-46753 Sequence 46753, A  
91 41 37.3 557 12 US-09-934-455-326 Sequence 326, App  
92 41 37.3 557 12 US-10-412-6998-870 Sequence 870, App  
93 41 37.3 557 12 US-10-423-066A-646 Sequence 646, App  
94 41 37.3 557 15 US-10-302-267-198 Sequence 198, App  
95 41 37.3 557 15 US-10-374-780A-2598 Sequence 2598, App  
96 41 37.3 639 12 US-09-920-954-4 Sequence 954, App  
97 41 37.3 660 12 US-10-425-114-64570 Sequence 64570, A  
98 41 37.3 933 9 US-09-815-242-10095 Sequence 10095, A  
99 41 37.3 933 9 US-09-815-242-10055 Sequence 10055, A  
100 41 37.3 933 12 US-10-282-122A-56482 Sequence 56482, A

## ALIGNMENTS

RESULT 1  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLIS, SCOTT E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/4210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 90.9%; Score 100; DB 9; Length 1375;  
Best Local Similarity 90.9%; Pred. No. 2,5e-07;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIRDI 22  
DB 578 VPSYSPARAHDSVQDLIRDI 599

RESULT 2  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLIS, SCOTT E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704

; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 90.0%; Score 99; DB 9; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 3.9e-07;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIRDI 22  
DB 552 VPSYSPARAHDSVQDLIRDI 573

RESULT 3  
US-09-995-749A-11  
; Sequence 11, Application US/09995749A  
; Patent No. US2002015568A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIKHUIZEN, LOUBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: B043388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patencin Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 522  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-995-749A-11

Query Match 71.8%; Score 79; DB 9; Length 522;  
Best Local Similarity 63.6%; Pred. No. 0.00023;  
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIRDI 22  
DB 146 IPYISFVRAHDSVQTLIAIV 167

RESULT 4  
US-09-995-749A-13  
; Sequence 13, Application US/09995749A  
; Patent No. US2002015568A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIKHUIZEN, LOUBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: B043388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; PRIOR FILING DATE: 2001-11-29

```

; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-13

Query Match
Best Local Similarity 61.8%; Score 68; DB 9; Length 535;
Pred. No. 0.015;
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIRDI 22
Db 144 IPNYSFVRAHDNNSQDQIQNAI 165

RESULT 5
US-09-995-749A-12
; Sequence 12, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-12

Query Match
Best Local Similarity 61.8%; Score 68; DB 9; Length 584;
Pred. No. 0.017;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIR 19
Db 167 IPNYSFVRAHDYDADPIR 185

RESULT 6
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
```

```

; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match
Best Local Similarity 61.8%; Score 68; DB 9; Length 1781;
Pred. No. 0.062;
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIRDI 22
Db 1123 IPNYSFVRAHDNNSQDQIQNAI 1144

RESULT 7
US-10-417-280A-2
; Sequence 2, Application US/10417280A
; Publication No. US2003022923A1
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: OLANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 0147-0247P
; CURRENT APPLICATION NUMBER: US/10/417,280A
; CURRENT FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: DE 19905069.4
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: US 09/499,203
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-10-417-280A-2

Query Match
Best Local Similarity 61.8%; Score 68; DB 15; Length 2057;
Pred. No. 0.073;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDLIR 19
Db 757 IPNYSFVRAHDYDADPIR 775

RESULT 8
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
```

US-09-995-749A-10

Query Match 56.4%; Score 62; DB 9; Length 545;  
Best Local Similarity 65.0%; Pred. No. 0.15;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARADSEVQDLIRDI 22  
Db 156 NYIFRAHDSVQTVIAKII 175

RESULT 9

US-09-740-274-6  
Sequence 6, Application US/09740274  
Patent No. US20020031826A1  
GENERAL INFORMATION:  
APPLICANT: Nicholas, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FaastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 56.4%; Score 62; DB 9; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.47;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARADSEVQDLIRDI 22  
Db 576 NYIFRAHDSVQTVIAKII 595

RESULT 10  
US-10-424-599-168435  
Sequence 168435, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 168435  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure

LOCATION: (1) ..(117)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_123111C.1.pep  
US-10-424-599-168435

Query Match 42.7%; Score 47; DB 12; Length 117;  
Best Local Similarity 45.0%; Pred. No. 7.5;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 3 SYSFARADSEVQDLIRDI 22  
Db 46 SYSYAMWHDSSVSTIPVDV 65

RESULT 11  
US-10-369-493-12332  
Sequence 12332, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 12332  
LENGTH: 462  
TYPE: PRT  
ORGANISM: Mesorhizobium loti  
US-10-369-493-12332

Query Match 42.7%; Score 47; DB 15; Length 462;  
Best Local Similarity 40.9%; Pred. No. 37;  
Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 1 VPSYFARADSEVQDLIRDI 22  
Db 290 VPSYHTAKFGRADLDAKIL 311

RESULT 12  
US-10-282-122A-76707  
Sequence 76707, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23

```
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 76707
;; LENGTH: 597
;; TYPE: PRT
;; ORGANISM: Ureaplasma urealyticum
US-10-282-122A-76707

Query Match      42.7%; Score 47; DB 12; Length 597;
Best Local Similarity 47.4%; Pred. No. 50;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      4 YSFARADSEVQDLIRDI 22
DB      173 YDPKRLNSELQRLIDSL 191

RESULT 13
US-10-282-122A-71700
;; Sequence 71700, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl
;; APPLICANT: Zyckind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: EITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
```

```
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 71700
;; LENGTH: 465
;; TYPE: PRT
;; ORGANISM: Staphylococcus haemolyticus
;; FEATURE:
;; NAME/KEY: MISC FEATURE
;; LOCATION: (11)-(11)
;; OTHER INFORMATION: X-any amino acid
US-10-282-122A-71700

Query Match      42.3%; Score 46.5; DB 12; Length 465;
Best Local Similarity 45.5%; Pred. No. 45;
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY      1 VPSYSPARADSEVQDLIR 19
DB      174 IPTYFPAVAIDHDYMEISDVIR 195

RESULT 14
US-10-424-599-232822
;; Sequence 232822, Application US/10424599
;; Publication No. US20040031072A1
;; GENERAL INFORMATION:
;; APPLICANT: La Rosa, Thomas J
;; APPLICANT: Kovalic, David K
;; APPLICANT: Zhou, Yinhua
;; APPLICANT: Cao, Yongwei
;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 232822
;; LENGTH: 54
;; TYPE: PRT
;; ORGANISM: Glycine max
;; FEATURE:
;; OTHER INFORMATION: Clone ID: PAT_MRT3847_52263C.1.pep
US-10-424-599-232822

Query Match      41.8%; Score 46; DB 12; Length 54;
Best Local Similarity 55.6%; Pred. No. 4.5;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY      1 VPSYSPARADSEVQDLIR 18
DB      10 VSSYSLARADHDHTGSKLL 27

RESULT 15
US-10-424-599-229462
;; Sequence 229462, Application US/10424599
;; Publication No. US20040031072A1
;; GENERAL INFORMATION:
;; APPLICANT: La Rosa, Thomas J
;; APPLICANT: Kovalic, David K
;; APPLICANT: Zhou, Yinhua
;; APPLICANT: Cao, Yongwei
;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 229462
;; LENGTH: 111
```

TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(111)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_49229C.1.pep  
US-10-424-599-229462

Query Match 41.8% Score 46; DB 12; Length 111;  
Best Local Similarity 40.0%; Pred. No. 10;  
Matches 10; Conservative 5; Mismatches 2; Indels 8; Gaps 1;

QY 1 VPSYSPARAH-----DSEVODL 17  
:|||||:|:|:|  
Db 85 LPSYSYHRAHATFCADYDIXIRDL 109

RESULT 16  
US-10-424-599-159618  
Sequence 159618, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 159618  
LENGTH: 170  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(170)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_115154C.1.pep  
US-10-424-599-159618

Query Match 41.8% Score 46; DB 12; Length 170;  
Best Local Similarity 53.8%; Pred. No. 17;  
Matches 7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 9 AHDSVODLIRDI 21  
:|||||:|:|:|  
Db 43 SNTDEVODLVKDL 55

RESULT 17  
US-09-815-242-5584  
Sequence 5584, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5584  
LENGTH: 481  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5584

Query Match 41.4% Score 45.5; DB 9; Length 481;  
Best Local Similarity 45.5%; Pred. No. 68;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSPARAHDS---EVODLIR 19  
:|||||:|:|:|  
Db 193 IPTYMFVAIDYMEISDVIR 214

RESULT 18  
US-09-815-242-12456  
Sequence 12456, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12456  
LENGTH: 487  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12456

Query Match 41.4% Score 45.5; DB 9; Length 487;  
Best Local Similarity 45.5%; Pred. No. 69;

Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSPARAHDS---EVODLIR 19  
Db 196 IFTNFAVAIDYMEISDVIR 217

## RESULT 19

US-10-424-599-264904  
; Sequence 264904, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 264904  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(166)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_81228C.1.pep  
US-10-424-599-264904

Query Match 40.9%; Score 45; DB 12; Length 166;  
Best Local Similarity 42.1%; Pred. No. 24;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 4 YSPARAHSEVODLIRDI 22  
Db 64 YLYAMVHDSVSSTIPRDFV 82

## RESULT 20

US-10-424-599-184218  
; Sequence 184218, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 184218  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_137365C.1.pep  
US-10-424-599-184218

Query Match 40.9%; Score 45; DB 12; Length 171;  
Best Local Similarity 42.1%; Pred. No. 25;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVODLIRDI 21  
Db 37 THSFSDHRSYDEDFLRDI 55

## RESULT 21

US-10-424-599-158664  
; Sequence 158664, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 158664  
; LENGTH: 362  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(362)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_114293C.1.pep  
US-10-424-599-158664

Query Match 40.9%; Score 45; DB 12; Length 362;  
Best Local Similarity 45.0%; Pred. No. 60;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 PSYFARAHSEVODLIRDI 21  
Db 10 PGMDMIMHDSRYDLVRDI 29

## RESULT 22

US-10-424-599-250160  
; Sequence 250160, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 250160  
; LENGTH: 365  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(365)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_67923C.1.pep  
US-10-424-599-250160

Query Match 40.9%; Score 45; DB 12; Length 365;  
Best Local Similarity 45.0%; Pred. No. 60;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 PSYFARAHSEVODLIRDI 21  
Db 10 PGMDMIMHDSRYDLVRDI 29



RESULT 23  
US-10-425-114-53178  
; Sequence 53178, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 53178  
; LENGTH: 377  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700840947\_FLI.pep  
US-10-425-114-53178

Query Match 40.9%; Score 45; DB 12; Length 377;  
Best Local Similarity 45.0%; Pred. No. 62;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 PSYSFARAHSEVODLRDI 21  
DB 27 PGMDMPIMHDSRDYDLVYDI 46

RESULT 24  
US-10-425-114-50726  
; Sequence 50726, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 50726  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700558525\_FLI.pep  
US-10-425-114-50726

Query Match 40.9%; Score 45; DB 12; Length 414;  
Best Local Similarity 45.0%; Pred. No. 70;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 PSYSFARAHSEVODLRDI 21  
DB 64 PGMDMPIMHDSRDYDLVYDI 83

RESULT 25  
US-10-425-114-65612  
; Sequence 65612, Application US/10425114  
; Publication No. US20040034888A1

; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 65612  
; LENGTH: 423  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700154223\_FLI.pep  
US-10-425-114-65612

Query Match 40.9%; Score 45; DB 12; Length 423;  
Best Local Similarity 42.9%; Pred. No. 71;  
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPSYFARAHSEVODLRDI 21  
DB 294 VPMYNTYRAHOGREARMRDM 314

RESULT 26  
US-10-424-599-194106  
; Sequence 194106, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic, David K  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 194106  
; LENGTH: 963  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(963)  
; OTHER INFORMATION: unsure at all Xaa locations  
; OTHER INFORMATION: Clone ID: PAT\_MFT3847\_17303C.1.pep  
US-10-424-599-194106

Query Match 40.9%; Score 45; DB 12; Length 963;  
Best Local Similarity 38.1%; Pred. No. 1,9e+02;  
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 PSYSFARAHSEVODLRDI 22  
DB 774 PXYLYAMVHDSVSSTIPGDFV 794

RESULT 27  
US-10-334-143-41  
; Sequence 41, Application US/10334143  
; Publication No. US20040009549A1  
; GENERAL INFORMATION:  
; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH  
; APPLICANT: SUDARSANAM, SUCHA

```

; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL
; FILE REFERENCE: KINASES IDENTIFIED WITH THE METHOD
; CURRENT APPLICATION NUMBER: US/10/334,143
; PRIOR FILING DATE: 2002-12-31
; PRIOR FILING DATE: 2001-12-31
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 2630
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-334-143-41

Query Match      40.9%; Score 45; DB 15; Length 2630;
Best Local Similarity 47.1%; Pred. No. 5.9e+02;
Matches      8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy      1 VPSYSPARAHDSSEVDLIR 17
Db      208 MPSTSLSGREEDL 224

RESULT 28
US-10-282-122A-71074
; Sequence 71074, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71074
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-10-282-122A-71074

Query Match      40.5%; Score 44.5; DB 12; Length 484;
Best Local Similarity 45.5%; Pred. No. 1e+02;
Matches      10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

Qy      1 VPSYSPARA---HDSSEVDLIR 19
Db      193 VPTNFAVAVDHHYQISDVIR 214

RESULT 29
US-10-424-599-267457
; Sequence 267457, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovacic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 267457
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(115)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_83537C.1.pep
; US-10-424-599-267457

Query Match      40.0%; Score 44; DB 12; Length 115;
Best Local Similarity 50.0%; Pred. No. 23;
Matches      9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy      3 SYSPARAHDSSEVDLIR 20
Db      49 SYXVALAHDSVSTIPGD 66

RESULT 30
US-09-862-027-37
; Sequence 37, Application US/09862027
; Patent No. US2002014248A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US2002014248A1 Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862,027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345,473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-862-027-37

Query Match      40.0%; Score 44; DB 9; Length 309;
Best Local Similarity 38.9%; Pred. No. 72;
Matches      7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy      5 SFPAHDSSEVDLIRDI 22
Db      190 SPEKVDPRIKIGSCI 207
```

Thu May 6 16:45:43 2004

us-09-290-049a-17.rapb

Page 10

Search completed: May 4, 2004, 09:25:40  
Job time : 36.3333 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:02:05 ; Search time 47.6667 Seconds  
(without alignments)  
130.407 Million cell updates/sec

Title: US-09-290-049a-18  
Perfect score: 110  
Sequence: 1 VPVYFIRAHDSVQTRIAKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : A\_Geneseq\_290a04:\*  
1: geneseqp19808:\*  
2: geneseqp19808:\*  
3: geneseqp20008:\*  
4: geneseqp20018:\*  
5: geneseqp20028:\*  
6: geneseqp20038:\*  
7: geneseqp20038:\*  
8: geneseqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1365	7	ADD93659 Streptococcus
2	92	83.6	1475	5	AAU98036 S. mutans
3	92	83.6	1475	5	AAU98037 S. mutans
4	92	83.6	1475	5	AAU98038 S. mutans
5	90	81.8	223	6	ABR63230 Glucanase
6	90	81.8	223	6	ABR63228 Glucanase
7	90	81.8	224	6	ABR63227 Glucanase
8	90	81.8	1430	5	AAU98043 S. mutans
9	90	81.8	1430	5	AAU98029 S. mutans
10	90	81.8	1430	7	ADD93656 Streptococcus
11	90	81.8	1497	6	ABR63234 Glucanase
12	90	81.8	2835	5	ABR98574 Dextranase
13	89	80.9	1527	5	AAU80055 Leucosin
14	89	80.9	1527	7	ADC54807 Streptococcus
15	87	79.1	1499	7	ADC54806 Protein S
16	87	79.1	1554	7	ADD93658 Streptococcus
17	86	78.2	1475	5	AAU98035 S. mutans
18	86	78.2	1475	5	AAU98034 S. mutans
19	86	78.2	1475	5	AAU98033 S. mutans
20	86	78.2	1475	5	AAU98032 S. mutans
21	84	76.4	1430	5	AAU98044 S. mutans
22	84	76.4	1430	5	AAU98045 S. mutans
23	84	76.4	1430	5	AAU98042 S. mutans
24	84	76.4	1430	5	AAU98041 S. mutans
25	83	75.5	2055	6	ABR63235 Glucanase

26	81	73.6	221	6	ABR63229	ABR63229 Glucanase
27	81	73.6	1149	5	ABR63236	ABR63236 Glucanase
28	80	72.7	1475	5	AAU98031	AAU98031 S. mutans
29	80	72.7	1475	5	AAU98040	AAU98040 S. mutans
30	80	72.7	1475	5	AAU98033	AAU98033 S. mutans
31	80	72.7	1475	5	AAU98030	AAU98030 S. mutans
32	80	72.7	1475	5	AAU98039	AAU98039 S. mutans
33	80	72.7	1475	5	AAU98027	AAU98027 S. mutans
34	80	72.7	1475	5	ADD93654	ADD93654 Streptococcus
35	77	70.0	1577	2	AAU91047	AAU91047 Alpha-D-G
36	76	69.1	15	5	ABR98652	ABR98652 Dextranase
37	76	69.1	1375	5	AAU98028	AAU98028 S. mutans
38	76	69.1	1375	5	AAU92828	AAU92828 Streptococcus
39	76	69.1	1375	7	ADD93655	ADD93655 Streptococcus
40	75	68.2	1017	5	AAU92985	AAU92985 Streptococcus
41	75	68.2	1476	5	AAU79284	AAU79284 Streptococcus
42	73	66.4	1518	7	ADD93660	ADD93660 Streptococcus
43	69	62.7	1590	7	ADD93657	ADD93657 Streptococcus
44	69	62.7	1592	2	AAU32925	AAU32925 Glucosylase
45	64	58.2	15	5	ABR98657	ABR98657 Dextranase
46	64	58.2	15	5	ABR98579	ABR98579 Dextranase
47	64	58.2	15	5	ABR98559	ABR98559 Peptide d
48	63	57.3	221	6	ABR63226	ABR63226 Glucanase
49	63	57.3	1781	5	AAU74519	AAU74519 Lactobacillus
50	63	57.3	2057	3	AAU10667	AAU10667 Lactobacillus
51	63	57.3	2147	6	ABR63231	ABR63231 Glucanase
52	62	56.4	15	5	ABR98654	ABR98654 Dextranase
53	61	55.5	15	5	ABR98655	ABR98655 Dextranase
54	61	55.5	15	5	ABR98653	ABR98653 Dextranase
55	61	55.5	2022	6	ABR63232	ABR63232 Glucanase
56	57	51.8	15	5	ABR98650	ABR98650 Dextranase
57	57	48.2	855	5	ABR98573	ABR98573 Dextranase
58	52	47.3	15	5	ABR98651	ABR98651 Dextranase
59	49	42.7	602	6	ABU48729	ABU48729 Protein e
60	47	42.7	336	6	ABU58754	ABU58754 Protein e
61	47	42.7	506	6	ABU19215	ABU19215 Protein e
62	47	42.7	557	6	ABU16175	ABU16175 Protein e
63	45.5	41.4	866	4	ABR64833	ABR64833 Drosophila
64	45	40.9	639	2	AAU17089	AAU17089 Arabidopsis
65	44	40.0	583	4	AAU64782	AAU64782 Arabidopsis
66	44	39.1	583	6	ABR72780	ABR72780 Arabidopsis
67	43	39.1	15	5	ABR98656	ABR98656 Dextranase
68	43	39.1	339	6	ABU20783	ABU20783 Protein e
69	43	39.1	339	6	ABU20782	ABU20782 Protein e
70	43	39.1	431	5	AAU37082	AAU37082 Amino acid
71	43	39.1	457	6	ABP55883	ABP55883 Bifidobacterium
72	43	38.6	319	5	ABR54548	ABR54548 Protein e
73	42.5	38.2	57	6	ABR80679	ABR80679 N. gonorrhoeae
74	42	38.2	57	6	ABR80186	ABR80186 N. gonorrhoeae
75	42	38.2	183	7	ADC95884	ADC95884 E. faecalis
76	42	38.2	346	6	ABM15844	ABM15844 Mycobacterium
77	42	38.2	442	4	ABR61572	ABR61572 Drosophila
78	42	38.2	484	6	ABU28399	ABU28399 Protein e
79	42	38.2	652	6	ADA34300	ADA34300 Acinetobacter
80	42	38.2	3118	4	AAU50362	AAU50362 Human SRC
81	42	37.3	78	4	AAU21090	AAU21090 Peptide #
82	41	37.3	78	4	ABR43405	ABR43405 Peptide #
83	41	37.3	78	4	AAU37289	AAU37289 Peptide #
84	41	37.3	78	4	ABR26377	ABR26377 Protein #
85	41	37.3	78	4	AAU77158	AAU77158 Human bone
86	41	37.3	416	6	AAU84536	AAU84536 Human bira
87	41	37.3	78	4	ABG58782	ABG58782 Human liv
88	41	37.3	78	4	ABG58782	ABG58782 Human pep
89	41	37.3	78	5	ABG46171	ABG46171 Supratena
90	41	37.3	98	6	ABP60161	ABP60161 Streptococcus
91	41	37.3	399	5	ABP29178	ABP29178 Streptococcus
92	41	37.3	416	6	AAU84536	AAU84536 Amino acid
93	41	37.3	457	6	ABU39823	ABU39823 Protein e
94	41	37.3	474	3	ABR64936	ABR64936 Drosophila
95	41	37.3	560	3	AAU42484	AAU42484 Human ORF
96	41	37.3	560	5	AAU17145	AAU17145 Human bio
97	41	37.3	560	5	AAU17145	AAU17145 Human bio
98	41	37.3	630	4	ABR58712	ABR58712 Drosophila

99 41 37.3 815 5 ABB93087 Abb93087 Herbicida  
100 41 37.3 961 6 ABG76429 Abg76429 Brome mos

## ALIGNMENTS

## RESULT 1

ADD93659 standard; protein; 1365 AA.

AC ADD93659;

DT 29-JAN-2004 (first entry)

DE Streptococcus downei glucosyltransferase-S.

KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

OS Streptococcus downei.

PN WO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006362.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402485P.

PA (FORS-) FORSYTH INST.

PI Smith DJ, Taubman MA;

DR WPI; 2003-845091/78.

PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

PS Claim 16; Page 15-16; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus downei  
XX glucosyltransferase-S (GTF-S). Peptide fragments of GTF-S, especially  
XX from the catalytic domain of the polypeptide, can be used in immunogenic  
XX compositions and subunit vaccines for dental caries. These compositions  
XX comprise a major histocompatibility complex (MHC) class II protein-  
XX binding peptide from S. mutants glucan binding protein-B (GbpB)  
XX covalently linked with a peptide fragment of a streptococcal  
XX glucosyltransferase. The compositions are used in a claimed method of  
XX eliciting production of an antibody in a mammal. Dieptopic or  
XX multipleptopic polypeptides can be prepared synthetically or by  
XX recombinant DNA technology. Antibodies raised against MHC class II  
XX binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1365 AA;

XX Query Match 100.0%; Score 110; DB 7; Length 1365;

XX Best Local Similarity 100.0%; Pred. No. 1.8e-10;

XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPMYVPIRAHDSVQTRIAKII 22

DB 537 VPMYVPIRAHDSVQTRIAKII 558

AAU98036 standard; protein; 1475 AA.

AAU98036;

DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant D567T/D571K.

DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

OS Streptococcus mutans.

OS Synthetic.

FN Key Location/Qualifiers

FT MISC-difference 567 /note= "Wild-type Asp substituted by Thr"

FT MISC-difference 571 /note= "Wild-type Asp substituted by Lys"

PN US2002031826-A1.

PD 14-MAR-2002.

PF 19-DEC-2000; 2000US-00740274.

PR 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

PA (NICH/) NICHOLS S E.

PI Nichols SE;

DR WPI; 2002-414332/44.

PT Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
XX B polypeptide having changes at position from I448V, D457N, D567T,  
XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D571K/K1014T, or a  
XX I448V/D457N/D567T/D571K/K1014T. Y169A/Y170A/Y171A, and K779Q or a  
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
XX GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
XX complementary polynucleotide, a ribonucleic acid sequence encoding the  
XX GTF mutant, an expression cassette comprising the polynucleotide operably  
XX linked to a promoter, a vector comprising the expression cassette, host  
XX cell introduced with the vector, a transgenic plant comprising the  
XX vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
XX coating composition comprising a glucan produced in a plant transformed  
XX with a gene encoding the mutant GTF, wild type or, starch, a latex.  
XX thermoplastic molecule or their combinations or glucan and starch where  
XX the glucan is produced in the amyloplast and/or vacuole or a maize line  
XX deficient in starch biosynthesis, transformed with a gene encoding a  
XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
XX comprising the glucan (paper sizing/coating agent). The vector is useful  
XX for producing a glucan in a plant. The method comprises transforming a  
XX plant cell with the vector, growing the plant cell under plant growing  
XX conditions to produce a regenerated plant and inducing expression of the  
XX polynucleotide for a time sufficient to produce the glucan in the  
XX regenerated plant, where the vector contains a transit sequence from  
XX ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
XX chlorophyll AB binding protein to produce a transgenic plant, and glucan  
XX is produced in the amyloplast of potato or the vacuole of sugar beet.  
XX Glucans are useful as substitutes for and additions to modified starch  
XX and latexes in paper manufacture. Unlike prior art techniques, which  
XX require input materials that produce chemical effluents, paper  
XX manufacture utilizing the glucan produced by GTF, which utilizes

CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTFB mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTFB sequence appearing as AAU98027 and the information in  
CC claim 36  
XX  
SQ Sequence 1475 AA;  
Query Match 83.6%; Score 92; DB 5; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 3,7e-07;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 VPNTVFIRADSEVQTRIAKII 22  
Db 552 VPSYSFIRADSEVQTRIAKII 573  
RESULT 3  
AAU98037 standard; protein; 1475 AA.  
XX  
AC AAU98037;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE 5. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.  
XX  
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
KW coating composition; glucan; starch; latex; thermoplastic molecule;  
KW amyloplast; vacuole; paper manufacture; mutant; mutain.  
XX  
OS Streptococcus mutans.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"  
FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"  
FT US2002031826-A1.  
XX  
PD 14-MAR-2002.  
XX  
PF 19-DEC-2000; 2000US-00740274.  
XX  
PR 07-JUN-1995; 95US-00478704.  
PR 07-JUN-1995; 95US-00482711.  
PR 07-JUN-1995; 95US-00485243.  
PR 16-JAN-1998; 98US-00007999.  
PR 16-JAN-1998; 98US-00008172.  
PR 20-JAN-1998; 98US-00009620.  
PR 11-DEC-1998; 98US-00210361.  
XX  
PA (NICH) NICHOLS S E.  
XX  
PI Nichols SE;  
XX  
DR WPI; 2002-414332/44.  
XX  
PT Glucosyltransferase B or D protein useful for producing a glucan useful  
PT as substitutes for and additions to modified starch and latexes in paper  
PT manufacture, comprises mutations in specific positions.  
XX  
PS Claim 36; Page; 44p; English.  
XX  
CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
CC B polypeptide having changes at position from 1448V, D457N, D567T,  
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,

CC 1448V/D457N/D567T/D571K/K790/K1014T, Y169A/Y170A/Y171A, and K790 or a  
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
CC GTF mutant, an expression cassette comprising the polynucleotide operably  
CC linked to a promoter, a vector comprising the expression cassette, host  
CC cell introduced with the vector, a transgenic plant comprising the  
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
CC coating composition comprising a glucan produced in a plant transformed  
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
CC thermoplastic molecule or their combinations or glucan and starch where  
CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTFB mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTFB sequence appearing as AAU98027 and the information in  
CC claim 36  
XX  
SQ Sequence 1475 AA;  
Query Match 83.6%; Score 92; DB 5; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 3,7e-07;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 VPNTVFIRADSEVQTRIAKII 22  
Db 552 VPSYSFIRADSEVQTRIAKII 573  
RESULT 4  
AAU98038 standard; protein; 1475 AA.  
XX  
AC AAU98038;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE 5. mutans GTFB mutant 1448V/D457N/D567T/D571K/K790/K1014T.  
XX  
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
KW coating composition; glucan; starch; latex; thermoplastic molecule;  
KW amyloplast; vacuole; paper manufacture; mutant; mutain.  
XX  
OS Streptococcus mutans.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 448 /note= "Wild-type Ile substituted by Val"  
FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"  
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"  
FT

FT Misc-difference 779  
 FT /note= "Wild-type Lys substituted by Gln"  
 FT Misc-difference 1014  
 FT /note= "Wild-type Lys substituted by Thr"  
 XX US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PD 19-DEC-2000; 2000US-00740274.  
 XX  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 XX (NICH/) NICHOLS S. E.  
 XX  
 XX Nichols SE;  
 XX  
 XX WPI; 2002-414332/44.  
 XX  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Claim 36; Page; 44pp; English.  
 XX  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC 1448V/D457N/D571K/K1014T, Y169A/Y170A/Y171A, and K790Q or a  
 CC GTF D polypeptide having changes at positions from 1589D, 1589E, N471D,  
 CC N471D/1589D, and N471D/1589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the expression cassette, host  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTF mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AA098027 and the information in  
 CC claim 36  
 XX  
 XX Sequence 1475 AA;  
 SQ

Query Match 83.6%; Score 92; DB 5; Length 1475;

Best Local Similarity 86.4%; Pred. No. 3.7e-07;  
 Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 VPNYVFIKRAHDSVQTRIAKII 22  
 DB 552 VPSYSPRIKRAHDSVQTRIAKII 573  
 RESULT 5  
 ABR63230  
 ID ABR63230 standard; protein; 223 AA.  
 XX  
 AC ABR63230;  
 XX  
 DT 27-AUG-2003 (first entry)  
 XX  
 DE Glucansucrase sequence from strain 1b86.  
 XX  
 KW Glucan; glucosyltransferase activity; thickener; prebiotic;  
 KW bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.  
 XX  
 XX Leuconostoc sp.  
 OS  
 PN WO2003008618-A2.  
 PD  
 PD 30-JAN-2003.  
 PF  
 PF 22-JUL-2002; 2002WO-NL000495.  
 XX  
 XX 20-JUL-2001; 2001EP-00202752.  
 PR 25-JUL-2001; 2001EP-00202841.  
 XX  
 XX (NEDS) NEDERLANDSE ORG TOEGEPAST.  
 PA  
 PA Van Geel- Schutten GH;  
 XX  
 XX WPI; 2003-289780/28.  
 DR  
 DR N-PSDB; ACC5074.  
 XX  
 XX Novel glucan produced by glucosyltransferase activity of lactic acid  
 PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.  
 PT  
 XX  
 PS Claim 11; Page 29; 51pp; English.  
 XX  
 XX The present invention relates to glucan capable of being produced by  
 CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
 CC substrate. The method is useful as a thickener, as a prebiotic and as a  
 CC bioactive agent and as an anti-corrosion agent. The glucan can be  
 CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
 CC products. The glucan is useful as anticorrosion agent, e.g. for the  
 CC protection of ship hulls. It can also be incorporated in nutritional or  
 CC pharmaceutical compositions intended for improving the condition of the  
 CC gastrointestinal tract. The present sequence represents a sequence of the  
 CC glucansucrase gene  
 CC  
 XX  
 XX Sequence 223 AA;  
 SQ

Query Match 81.8%; Score 90; DB 6; Length 223;  
 Best Local Similarity 85.7%; Pred. No. 8.7e-08;  
 Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 PNYVFIKRAHDSVQTRIAKII 22  
 DB 141 PNYVFIKRAHDSVQTRIAKII 161  
 RESULT 6  
 ABR63228  
 ID ABR63228 standard; protein; 223 AA.  
 XX  
 AC ABR63228;  
 XX

DT	27-AUG-2003	(first entry)
XX		
DE	Glucanase sequence from strain LB86.	
XX		
KM	Glucan; glucosyltransferase activity; thickener; prebiotic;	
XX	bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.	
OS	Leuconostoc sp.	
XX		
FN	MO2003008618-A2.	
XX		
PD	30-JAN-2003.	
XX		
FE	22-JUL-2002; 2002WO-NL000495.	
XX		
PR	20-JUL-2001; 2001EP-00202752.	
XX		
PR	25-JUL-2001; 2001EP-00202841.	
XX		
PA	(NED) NEDERLANDSE ORG TOEGEPAST.	
PI	Van Geel - Schutten GH;	
XX		
DR	WP1; 2003-289780/28.	
DR	N-PADB; ACC50072.	
XX		
PT	Novel glucan produced by glucosyltransferase activity of lactic acid	
PT	bacterium on sucrose substrate, and having backbone consisting of alpha	
PT	(1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.	
BS	Claim 11; Page 27; 51pp; English.	
XX		
CC	The present invention relates to glucan capable of being produced by	
CC	glucosyltransferase activity of a lactic acid bacterium on a sucrose	
CC	substrate. The method is useful as a thickener, as a prebiotic and as a	
CC	bioactive agent and as an anti-corrosion agent. The glucan can be	
CC	incorporated in foodstuffs such as beverages, sauces, dressings, dairy	
CC	products. The glucan is useful as anticorrosion agent, e.g. for the	
CC	protection of ship hulls. It can also be incorporated in nutritional or	
CC	pharmaceutical compositions intended for improving the condition of the	
CC	gastrointestinal tract. The present sequence represents a sequence of the	
CC	glucanase gene	
XX		
SQ	Sequence 223 AA;	
XX		
Query Match	81.8%; Score 90; DB 6; Length 223;	
Better Local Similarity	77.3%; Pred. No. 8.7e-08;	
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;		
OY	1 VENVVFIRAHDSVOTRIAKIT 22	
	:   :       :    :	
DB	140 IPNYSFVRHDSVQTIVIAIIL 161	
RESULT 7		
ID	ABR63227 standard; protein; 224 AA.	
XX		
AC	ABR63227;	
XX		
DT	27-AUG-2003 (first entry)	
DE	Glucanase sequence from strain LB33.	
XX		
KM	Glucan; glucosyltransferase activity; thickener; prebiotic;	
XX	bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.	
OS	Lactobacillus sp.	
XX		
PN	MO2003008618-A2.	
XX		
PD	30-JAN-2003.	
XX		
PF	22-JUL-2002; 2002WO-NL000495.	

Query Match	81.8%	Score 90	DB 6	Length 224
Best Local Similarity	85.7%	Pred. No. 8.7e-08		
Matches 18	Conservative 1	Mismatches 2	Indels 0	Gaps 0
Qy	2 PNYVFIRAHDSKVOTRIAKII 22			
Db	142 PNYTFIRAHDSKVOTRIAKII 162			
XX	20-JUL-2001; 2001EP-00202752.			
PR	25-JUL-2001; 2001EP-00202841.			
XX	(NEDE ) NEDERLANDSE ORG TOEGEPAST.			
PA	Van Geel- Schulten GH;			
PI	WPI; 2003-289780/28.			
XX	N-PSDB; ACC50071.			
XX	Novel glucan produced by glucosyltransferase activity of lactic acid			
FT	bacterium on sucrose substrate, and having backbone consisting of alpha			
PR	(1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.			
XX	Claim 11; Page 25-26; 51pp; English.			
XX	The present invention relates to glucan capable of being produced by			
CC	glucosyltransferase activity of a lactic acid bacterium on a sucrose			
CC	substrate. The method is useful as a thickener, as a prebiotic and as a			
CC	biocative agent and as an anti-corrosion agent. The glucan can be			
CC	incorporated in foodstuffs such as beverages, sauces, dressings, dairy			
CC	products. The glucan is useful as anticorrosion agent, e.g. for the			
CC	protection of ship hulls. It can also be incorporated in nutritional or			
CC	pharmaceutical compositions intended for improving the condition of the			
CC	gastrointestinal tract. The present sequence represents a sequence of the			
CC	glucanucrase gene			
XX	Sequence 224 AA;			
XX	RESULT 8			
XX	AAU98043			
XX	ID AAU98043 standard; protein; 1430 AA.			
XX	AAU98043;			
DT	27-AUG-2002 (first entry)			
XX	S. mutans glucosyltransferase GTFD mutant N471D.			
XX	Glucosyltransferase; GTFD; transgenic plant; paper sizing;			
KW	coating composition; glucan; starch; latex; thermoplastic molecule;			
KW	amyloplastic; vacuole; paper manufacture; mutant; mutain.			
XX	Streptococcus mutans.			
OS	Synthetic.			
XX	Key	Location/Qualifiers		
FT	Misc-difference 471	/note= "Wild-type Asn substituted by Asp"		
XX	US2002031826-A1.			
XX	14-MAR-2002.			
PD	19-DEC-2000; 2000US-00740274.			
XX	07-JUN-1995; 95US-00478704.			
PR	07-JUN-1995; 95US-00482711.			
PR	16-JAN-1998; 98US-000485243.			
PR	16-JAN-1998; 98US-00008172.			
PR	20-JAN-1998; 98US-00009620.			
PR	11-DEC-1998; 98US-00210361.			



XX  
PA (NICH/) NICHOLS S E.

PI Nichols SE;

DR WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44pp; English.

CC The invention an isolated protein

CC K1014T, D457N/D567T, D457N/D567T  
CC T448V/D467N/D567T/D571K/K779Q

The invention an isolated protein comprising a glucosyltransferase (GTF), B polypeptide having changes at position from 1448Y, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K/D571K/K1014T, 1448Y/D457N/D567T/D571K/K179Q/K1014T, Y169A/Y170A/Y171A, GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations and/or vacuole or a male line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, paper require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTF mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexex using the GTFD sequence appearing as MU98029 and the information in claim 36

Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQVRIAKII 22  
 ||:|||||  
 DB 576 NYIFIRAHDSVQVRIAKII 595

## RESULT 10

ADD93656  
 ID ADD93656 standard; protein; 1430 AA.

AC ADD93656;

DT 29-JAN-2004 (first entry)

DE Streptococcus mutans glucosyltransferase-D.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX WO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006962.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

PS Claim 16; Page 13-14; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus mutans glucosyltransferase-D (GTF-D). Peptide fragments of GTF-D, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1430 AA;

Query Match 81.8%; Score 90; DB 7; Length 1430;  
 Best Local Similarity 90.0%; Pred. No. 8.2e-07;  
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQVRIAKII 22  
 ||:|||||  
 DB 576 NYIFIRAHDSVQVRIAKII 595

## RESULT 11

ABR63234  
 ID ABR63234 standard; protein; 1497 AA.

XX ABR63234;

XX 27-AUG-2003 (first entry)

XX Glucanase sequence from strain Ib33.

DE Glucan; glucosyltransferase activity; thickener; prebiotic;

XX bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

XX Lactobacillus sp.

XX WO2003008618-A2.

XX 30-JAN-2003.

XX 22-JUL-2002; 2002WO-NL000495.

XX 20-JUL-2001; 2001EP-00202752.

XX 25-JUL-2001; 2001EP-00202841.

XX (NED) NEDERLANDSE ORG TORGERAST.

XX Van Geel- Schutten GH;

XX WPI; 2003-289780/28.

XX N-PSDB; ACC84450.

XX Novel glucan produced by glucosyltransferase activity of lactic acid bacterium on sucrose substrate, and having backbone consisting of alpha (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.

XX Claim 11; Fig 1; 51pp; English.

XX The present invention relates to glucan capable of being produced by glucosyltransferase activity of a lactic acid bacterium on a sucrose substrate. The method is useful as a thickener, as a prebiotic and as a bioactive agent and as an anti-corrosion agent. The glucan can be incorporated in foodstuffs such as beverages, sauces, dressings, dairy products. The glucan is useful as anticorrosion agent, e.g. for the protection of ship hulls. It can also be incorporated in nutritional or pharmaceutical compositions intended for improving the condition of the gastrointestinal tract. The present sequence represents a sequence of the glucanase gene

XX Sequence 1497 AA;

Query Match 81.8%; Score 90; DB 6; Length 1497;  
 Best Local Similarity 85.7%; Pred. No. 8.7e-07;  
 Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 PNVFIKRAHDSVQVRIAKII 22  
 ||:|||||  
 DB 605 PNVFIKRAHDSVQVRIAKII 625

ABR98574  
 ID ABR98574 standard; protein; 2835 AA.

XX ABR98574;

XX 29-AUG-2003 (revised)

DT 14-JAN-2003 (first entry)

XX Dextran saccharase, DSR.

XX Dextran saccharase; enzyme; cytotactic; dermatological; antiseborrheic;

XX DSR-B; glycosyl transferase; dextran; prebiotic; pharmaceutical;

XX microflora regulation; intestinal transit; mineral assimilation;

XX colon cancer; acne; dandruff; body odour.

XX Leuconostoc mesenteroides; NRRL B-1299.

XX Key Location/Qualifiers  
 FT Peptide 1..40 /label=Signal\_peptide

FT	Protein	41..2835
FT	/label= Mature_protein	
FT	Peptide	423..439
FT	/label= SEQ_ID_6	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	478..501
FT	/label= SEQ_ID_7	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	519..539
FT	/label= SEQ_ID_8	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	560..571
FT	/label= SEQ_ID_9	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	631..645
FT	/label= SEQ_ID_10	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	1014..1021
FT	/label= SEQ_ID_11	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Domain	1981..1142
FT	/note= "Catalytic domain, SEQ ID 1. This sequence is specifically claimed in Claim 4"	
FT	Peptide	2120..2138
FT	/label= SEQ_ID_12	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	2161..2184
FT	/label= SEQ_ID_13	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	2202..2214
FT	/label= SEQ_ID_14	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	2243..2250
FT	/label= SEQ_ID_15	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	2315..2322
FT	/label= SEQ_ID_16	
FT	/note= "This sequence is specifically claimed in Claim 7"	
FT	Peptide	2689..2696
FT	/label= SEQ_ID_17	
FT	/note= "This sequence is specifically claimed in Claim 7"	
XX	FR2822163-A1.	
XX	20-SEP-2002.	
XX	19-DEC-2001, 2001PR-00016495.	
XX	16-MAR-2001, 2001PR-00003631.	
XX	(CNRS ) CNRS CENT NAT RECH SCT.	
XX	Bozonnet SAM, Remaud SMMC, Willemot RML, Moneau PEF,	
XX	WPI, 2002-715213/78.	
XX	N-PDSB; ABQ80961, ABQ80962.	
XX	New glycosyl transferase enzymes, containing glucan bonding and catalytic domains and producing alpha-(1-2) branched dextrans, useful in probiotic, pharmaceutical or cosmetic compositions.	
XX	Claim 6; Page 65-74; 82pp; French.	
CC	The present sequence is a novel dextran saccharase, DSR-E, from	
CC	Leuconostoc mesenteroides NRRL B-1299. The saccharase has glycosyl	
CC	transferase activity suitable for producing dextrans having alpha(1-2)	
CC	branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside, alpha-	
CC	fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbiopyranoside or alpha-D-	
CC	galactopyranosyl-sucrose. The dextran saccharase is useful in probiotic,	
CC	pharmaceutical or cosmetic compositions. The dextrans and related	
CC	compounds having alpha(1-2) bonds, produced using DSR-E, may be involved	
CC	in signalling/cellular recognition processes in vivo (specifically in	
CC	regulation of microflora in the intestines or on the skin); and are	

CC	potentially useful for improving intestinal transit, increasing assimilation of minerals (e.g. calcium and/or magnesium), preventing cancer of the colon and combating skin problems such as acne, dandruff and body odour. (Updated on 29-AUG-2003 to standardise OS field)
XX	
XX	
SQ	Sequence 2835 AA;
Query Match	81.8%; Score 90; DB 5; Length 2835;
Best Local Similarity	85.7%; Pred. No. 1.9e-06;
Matches 18; Conservative 1;	Mismatches 2; Indels 0; Gaps 0;
OY	2 PNVFIRAHDSVQTRIAKII 22               :   PNTAFIRAHDSVQTVIAQII 649
Db	629 PNTAFIRAHDSVQTVIAQII 649
RESULT 13	
ID	ABR55594 standard; protein; 2835 AA.
XX	ABR55594;
AC	
XX	
DT	11-AUG-2003 (first entry)
XX	
DE	Amino acid sequence of a glycosyltransferase designated DSR-D.
KW	Glycosyltransferase; DSR-D; enzyme; dextran; sucrose;
KW	alpha-D-flucoroglucose; p-nitrophenol alpha-D-glucopyranoside;
KW	alpha-D-glucopyranoside-alpha-D-sorbitofuranoside; oligosaccharide;
KW	oligoside; 4-O-alpha-D-galactopyranosylsucrose; weight gain;
KM	colon cancer; skin disorder; acne; dandruff; body odour;
KM	intestinal transit; calcium; magnesium.
OS	Leuconostoc mesenteroides.
XX	
PN	FR2822162-Al.
XX	
PD	20-SEP-2002.
XX	
PF	16-MAR-2001; 2001FR-00003631.
XX	
PR	16-MAR-2001; 2001FR-00003631.
PA	(MAG-) INST NAT SCI APLIQUERS TOULOUSE. (CNRS ) CNRS CENT NAT RECH SCT.
XX	
PI	Monnan P;
XX	
DR	WPI; 2003-432740/41.
XX	N-PsDB; ACC70331.
PT	New polypeptide with glycosyltransferase activity, useful for producing dextrans with alpha 1-2 branches, e.g. used as prebiotics.
PS	Claim 6; Fig 8; 103pp; French.
XX	The present sequence represents a glycosyltransferase, designated DSR-D.
CC	This enzyme produces dextrans having alpha(1-2) branches from sucrose,
CC	alpha-D-flucoroglucose, p-nitrophenol alpha-D-glucopyranoside, alpha-D-
CC	glucopyranoside-alpha-D-sorbitofuranoside or 4-O-alpha-D-
CC	galactopyranosylsucrose. The polypeptide is used to prepare
CC	oligosaccharides and oligosides for prebiotic, pharmaceutical, diagnostic
CC	and cosmetic applications, e.g. to improve health and weight gain in
CC	animals; prevent cancer of the colon; and treat skin disorders such as
CC	acne, dandruff and body odour. The oligosaccharides and oligosides
CC	produced by DSR-D improve intestinal transit and assimilation of calcium,
CC	magnesium and other minerals
XX	
SQ	Sequence 2835 AA;
Query Match	81.8%; Score 90; DB 6; Length 2835;
Best Local Similarity	85.7%; Pred. No. 1.9e-06;
Matches 18; Conservative 1;	Mismatches 2; Indels 0; Gaps 0

QY 2 PNYVFIRAHDSVQTRIAKII 22  
 DB 629 PNYVFIRAHDSVQTRIAQII 649

RESULT 14  
 AAU80055  
 ID AAU80055 standard; protein; 1527 AA.  
 AC AAU80055;  
 XX 30-JUL-2002 (first entry)

DE Leuconostoc mesenteroides dextran sucrose.  
 DE Dextran sucrose; yoghurt; curd; cheese; fermented milk; infant formulae;  
 KW pet food; vitamin; oral vaccine; enzyme.  
 XX Leuconostoc mesenteroides.  
 OS EPI201131-A1.  
 XX EPI201131-A1.  
 XX 02-MAY-2002.  
 XX 23-OCT-2000; 2000EP-00123012.  
 XX 23-OCT-2000; 2000EP-00123012.  
 XX (NEST ) SOC PROD NESTLE SA.  
 PA Bauche A, De Maleprade D, Duboc P, Neubauer H, Zink R;  
 PI WPI; 2002-373873/41.  
 DR N-PSDB; ABKS0932.

PT Preparing fermented food products, e.g. yoghurt, using microorganisms  
 PT that have dextran sucrose bound to the surface, provide attractive and  
 PT uniform texture.  
 XX  
 PS Disclosure; Page 14-20; 38pp; English.

XX The invention relates to preparation of a fermented food product that  
 CC includes creating the microorganism (A) to be used for fermentation with  
 CC a dextran sucrose so that this binds to the cell walls of (A). The method  
 CC is used to produce yoghurt, curd, cheese or other fermented milk  
 CC products, i.e. cream, fermented cereal products, water-based jellies,  
 CC infant formulae and pet foods. Dextran sucrose can also be bound to e.g.  
 CC vitamins or oral vaccines for delivering these to foods, using (A) as  
 CC carrier. The method produces foods with attractive and uniform texture,  
 CC since dextran sucrose is evenly distributed and forms a thickening agent  
 CC in situ, with the amount formed being controlled by the amount of  
 CC dextran sucrose bound. Milk does not have to be coagulated in order to  
 CC achieve a good texture, so the amount of acidic by-products formed is  
 CC reduced. Dextran sucrose binds to many different types of cells over wide  
 CC ranges of temperature and pH and the treated bacteria can be stored in  
 CC liquid or powdered form. The present sequence represents the amino acid  
 CC sequence of dextran sucrose

XX Sequence 1527 AA;

QY Query Match 80.9%; Score 89; DB 5; Length 1527;  
 Best Local Similarity 72.7%; Pred. No. 1.4e-06;  
 Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNYVFIRAHDSVQTRIAKII 22  
 DB 652 IPNYSFVRAHDSVQTRIAQIV 673

RESULT 15  
 ADC54807  
 ID ADC54807 standard; protein; 1527 AA.

XX ADC54807;  
 AC 18-DEC-2003 (first entry)

DE Leuconostoc mesenteroides dextran sucrose protein.  
 DE Dextran sucrose; active centre zone; glucan; polysaccharide; dextran;  
 KW D-glucose; starch; cellulose; glucan manufacture; transduction;  
 KW enzyme-reaction product.  
 XX Leuconostoc mesenteroides.  
 OS JP2003111590-A.  
 XX JP2003111590-A.  
 XX 15-APR-2003.  
 XX 03-OCT-2001; 2001JP-00307067.  
 XX 03-OCT-2001; 2001JP-00307067.  
 XX (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUHIN SOGO KEN.  
 PA WPI; 2003-735670/70.  
 DR N-PSDB; ADC54802.

PT Novel modified dextran sucrose which exchanges one site of active center  
 PT zone of dextran sucrose for active center zone of different types of  
 PT dextran sucrose, useful for manufacturing glucan.  
 XX  
 PS Example 1; SEQ ID NO 12; 28pp; Japanese.

XX This invention relates to a modified dextran sucrose (DS) exchanging one  
 CC site of the active centre zone of a dextran sucrose for the active centre  
 CC zone of different types of dextran sucrose. Glucan is a polysaccharide  
 CC (for example dextran) which uses D-glucose, such as a starch and a  
 CC cellulose, as a structural unit. The modified enzyme of the invention is  
 CC useful in the manufacture of glucan. The modified enzyme of the active centre  
 CC area of DS which carries out transduction, enables changes in structure  
 CC and character of an enzyme-reaction product and their application to  
 CC various uses. The present sequence is that of the Leuconostoc  
 CC mesenteroides dextran sucrose protein used during the exemplification of  
 CC the invention.

XX Sequence 1527 AA;

QY Query Match 80.9%; Score 89; DB 7; Length 1527;  
 Best Local Similarity 72.7%; Pred. No. 1.4e-06;  
 Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNYVFIRAHDSVQTRIAKII 22  
 DB 652 IPNYSFVRAHDSVQTRIAQIV 673

RESULT 16  
 ADC54806  
 ID ADC54806 standard; protein; 1499 AA.

XX ADC54806;  
 XX 18-DEC-2003 (first entry)

DE Protein Seq ID11 related to L mesenteroides dextran sucrose protein.  
 DE Dextran sucrose; active centre zone; glucan; polysaccharide; dextran;  
 KW D-glucose; starch; cellulose; glucan manufacture; transduction;  
 KW enzyme-reaction product.  
 XX Unidentified.  
 OS JP2003111590-A.  
 XX

PD 15-APR-2003.  
 XX  
 CC 03-OCT-2001; 2001JP-00307067.  
 XX  
 CC 03-OCT-2001; 2001JP-00307067.  
 XX  
 CC (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUHIN SOGO KEN.  
 XX  
 CC WPI; 2003-735670/70.  
 DR P-PSDB; ADC54814.  
 XX  
 CC Novel modified dextran sucrose which exchanges one site of active center  
 PT zone of dextran sucrose for active center zone of different types of  
 PT dextran sucrose, useful for manufacturing glucan.  
 XX  
 PS Example 1; SEQ ID NO 11; 28pp; Japanese.  
 CC  
 CC This invention relates to a modified dextran sucrose (DS) exchanging one  
 CC site of the active centre zone of a dextran sucrose for the active centre  
 CC zone of different types of dextran sucrose. Glucan is a polysaccharide  
 CC (for example dextran) which uses D-glucose, such as a starch and a  
 CC cellulose, as a structural unit. The modified enzyme of the active centre  
 CC is useful in the manufacture of glucan. The selection of the active centre  
 CC area of DS which carries out transduction, enables changes in structure  
 CC and character of an enzyme-reaction product and their application to  
 CC various uses. The present sequence is that of a protein which is related  
 CC to the leuconostoc mesenteroides dextran sucrose protein and which was  
 CC used during the exemplification of the invention.  
 CC  
 SQ Sequence 1499 AA;  
 XX  
 XX Query Match 79.1%; Score 87; DB 7; Length 1499;  
 XX Best Local Similarity 81.0%; Pred. No. 3.1e-06;  
 XX Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 PNVTFFIRAHDSFVQTRIAXII 22  
 DB 625 PNVSFVRADHSEVQTVIAXII 645  
 RESULT 17  
 ID ADD93658 standard; protein; 1554 AA.  
 XX  
 AC ADD93658;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus sobrinus glucosyltransferase-U.  
 XX  
 XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
 XX  
 OS Streptococcus sobrinus.  
 XX  
 PN W02003075845-A2.  
 PD 18-SEP-2003.  
 XX  
 PF 07-MAR-2003; 2003WO-US006962.  
 XX  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX  
 PA (FORS-) FORSYTH INST.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 DR WPI; 2003-845091/78.  
 XX  
 PT Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX

PS Claim 16; Page 15; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of Streptococcus sobrinus  
 CC glucosyltransferase-U (GTF-U). Peptide fragments of GTF-U, especially  
 CC from the catalytic domain of the polypeptide, can be used in immunogenic  
 CC compositions and subunit vaccines for dental caries. These compositions  
 CC comprise a major histocompatibility complex (MHC) class II protein-  
 CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
 CC covalently linked with a peptide fragment of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Dieptopic or  
 CC multiepitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 CC  
 SQ Sequence 1554 AA;  
 XX  
 XX Query Match 79.1%; Score 87; DB 7; Length 1554;  
 XX Best Local Similarity 72.7%; Pred. No. 3.2e-06;  
 XX Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 PNVTFFIRAHDSFVQTRIAXII 22  
 DB 557 IPTVSFVRADHSEVQTVIAXIV 578  
 RESULT 18  
 ID AAU98035 standard; protein; 1475 AA.  
 XX  
 AC AAU98035;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant D457N/D571K.  
 XX  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX amyloplast; vacuole; paper manufacture; mutant; muten.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"  
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"  
 FT  
 XX  
 XX US2002031826-A1.  
 XX  
 PN 14-MAR-2002.  
 PD  
 PD 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 XX (NICH/) NICHOLS S B.  
 XX  
 PA Nichols SE;  
 XX  
 PI  
 XX  
 DR WPI; 2002-414332/44.  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX

PS Claim 36; Page: 44pp; English.

CC The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from 1448V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,

CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line

CC deficient in starch biosynthesis, transformed with a gene encoding a

CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the

CC regenerated plant, where the vector contains a transit sequence from

CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper

CC manufacture utilizing the glucan produced by GTF, which utilizes

CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTFB mutant of the invention. Note: The present

CC sequence is not shown in the specification but was created by the indexer

CC using the GTFB sequence appearing as AAU98027 and the information in

CC claim 36

CC XX

SO Sequence 1475 AA;

Query Match 78.2%; Score 86; DB 5; Length 1475;

Best Local Similarity 81.8%; Pred. No. 4.6e-06;

Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPMVVFIRADSEVQRIAKII 22

DB 552 VPSYSTRADSEVQDIKII 573

RESULT 19

AAU98034

ID AAU98034 standard; protein: 1475 AA.

AC AAU98034;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant D457N/D567T.

XX

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

XX coating composition; glucan; starch; latex; thermoplastic molecule;

XX amyloplast; vacuole; paper manufacture; mutant; mutain.

XX

OS Streptococcus mutans.

OS Synthetic.

XX

XX Key Location/Qualifiers

XX Misc-difference 457

XX /note= "Wild-type Asp substituted by Asn"

XX

XX Misc-difference 567

FT /note= "Wild-type Asp substituted by Thr"

XX

XX US2002031826-A1.

XX

XX 14-MAR-2002.

XX

XX 19-DEC-2000; 2000US-00740274.

XX

XX 07-JUN-1995; 98US-00478704.

XX

XX 07-JUN-1995; 95US-00482711.

XX

XX 07-JUN-1995; 95US-00485243.

XX

XX 16-JAN-1998; 98US-00007999.

XX

XX 16-JAN-1998; 98US-00008172.

XX

XX 20-JAN-1998; 98US-00009620.

XX

XX 11-DEC-1998; 98US-00210361.

XX

XX (NICH/) NICHOLS S E.

XX

XX Nichols SE;

XX

XX WPI, 2002-414332/44.

XX

XX Glucosyltransferase B or D protein useful for producing a glucan useful

XX as substitutes for and additions to modified starch and latexes in paper

XX manufacture, comprises mutations in specific positions.

XX

XX Claim 36; Page: 44pp; English.

XX

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

XX B polypeptide having changes at position from 1448V, D457N, D567T,

XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,

XX I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,

XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

XX GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

XX complementary polynucleotide, a ribonucleic acid sequence encoding the

XX GTF mutant, an expression cassette comprising the polynucleotide operably

XX linked to a promoter, a vector comprising the expression cassette, host

XX cell introduced with the vector, a transgenic plant comprising the

XX vector, a seed or tuber from the transgenic plant, a paper sizing and/or

XX coating composition comprising a glucan produced in a plant transformed

XX with a gene encoding the mutant GTF, wild type or, starch, a latex,

XX thermoplastic molecule or their combinations or glucan and starch where

XX the glucan is produced in the amyloplast and/or vacuole or a maize line

XX deficient in starch biosynthesis, transformed with a gene encoding a

XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper

XX comprising the glucan (paper sizing/coating agent). The vector is useful

XX for producing a glucan in a plant. The method comprises transforming a

XX plant cell with the vector, growing the plant cell under plant growing

XX conditions to produce a regenerated plant and inducing expression of the

XX polynucleotide for a time sufficient to produce the glucan in the

XX regenerated plant, where the vector contains a transit sequence from

XX ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

XX chlorophyll AB binding protein to produce a transgenic plant, and glucan

XX is produced in the amyloplast of potato or the vacuole of sugar beet.

XX Glucans are useful as substitutes for and additions to modified starch

XX and latexes in paper manufacture. Unlike prior art techniques, which

XX require input materials that produce chemical effluents, paper

XX manufacture utilizing the glucan produced by GTF, which utilizes

XX biologically produced input materials, is more cost-effective and

XX environmentally friendly. Moreover, glucans also exhibit thermoplastic

XX properties and impart gloss to the paper during coating step. The present

XX sequence represents a GTFB mutant of the invention. Note: The present

XX sequence is not shown in the specification but was created by the indexer

XX using the GTFB sequence appearing as AAU98027 and the information in

XX claim 36

XX

SO Sequence 1475 AA;

Query Match 78.2%; Score 86; DB 5; Length 1475;

Best Local Similarity 81.8%; Pred. No. 4.6e-06;

Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VENTVFIRADSEVQTRIAKII 22  
 ||:|||||  
 Db 552 VPSYFIRADSEVQTRIADII 573

RESULT 20  
 AAU98032  
 ID AAU98032 standard; protein, 1475 AA.  
 AC AAU98032;  
 DT 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB mutant D567T.  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 OS Key Location/Qualifiers  
 FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
 FT US2002031826-A1.  
 XX 14-MAR-2002.  
 PF 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 PA Nichols SE;  
 PI WPI; 2002-414332/44.  
 DR WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, and K779Q or a  
 CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y163A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the expression cassette, host  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant, a paper sizing and/or  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

SQ Sequence 1475 AA;

Query Match 78.2%; Score 86; DB 5; Length 1475;  
 Best Local Similarity 81.8%; Pred. No. 4,6e-06;  
 Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VENTVFIRADSEVQTRIAKII 22  
 ||:|||||  
 Db 552 VPSYFIRADSEVQTRIADII 573

RESULT 21  
 AAU98044  
 ID AAU98044 standard; protein, 1430 AA.  
 AC AAU98044;  
 DT 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFD mutant N471D/T589D.  
 XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 OS Key Location/Qualifiers  
 FT Misc-difference 471 /note= "Wild-type Asn substituted by Asp"  
 FT FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"  
 FT US2002031826-A1.  
 XX 14-MAR-2002.  
 PD 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 PA Nichols SE;  
 PI WPI; 2002-414332/44.  
 DR WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful



PT as substitutes for and additions to modified starch and latexes in paper  
PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
CC B polypeptide having changes at position from 1448V, D457N, D567T,  
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
CC 1448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
CC GTF mutant, an expression cassette comprising the polynucleotide operably  
CC linked to a promoter, a vector comprising the expression cassette, host  
CC cell introduced with the vector, a transgenic plant comprising the  
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
CC coating composition comprising a glucan produced in a plant transformed  
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTF mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTF sequence appearing as AAU98029 and the information in  
CC claim 36

XX Sequence 1430 AA;

Query Match 76.4%; Score 84; DB 5; Length 1430;  
Best Local Similarity 85.0%; Pred. No. 1e-05; Mismatches 0; Gaps 0;  
Matches 17; Conservative 1; Indels 2;

QY 3 NYVPIRAHSEVQDTRAKII 22  
DB 576 NYVPIRAHSEVQDTRAKII 595

RESULT 22

AAU98045 standard; protein; 1430 AA.

XX AAU98045;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD mutant N471D/T589E.

XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
KM coating composition; glucan; starch; latex; thermoplastic molecule;  
KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Location/Qualifiers

XX Key

PT Misc-difference 471 /note= "Wild-type Asn substituted by Asp"  
FT FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"  
PT PT  
XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
CC B polypeptide having changes at position from 1448V, D457N, D567T,  
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
CC 1448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
CC GTF mutant, an expression cassette comprising the polynucleotide operably  
CC linked to a promoter, a vector comprising the expression cassette, host  
CC cell introduced with the vector, a transgenic plant comprising the  
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
CC coating composition comprising a glucan produced in a plant transformed  
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTF mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTF sequence appearing as AAU98029 and the information in  
CC claim 36

XX Sequence 1430 AA;

Query Match 76.4%; Score 84; DB 5; Length 1430;



Best Local Similarity 85.0%; Pred. No. 1e-05;  
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYVFIKRAHSEVOGTRIAKII 22  
||:|||||||  
Db 576 NYVFIKRAHSEVOGTRIAKII 595

RESULT 23  
AAU98042

ID AAU98042 standard; protein; 1430 AA.

AC AAU98042;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFD mutant T589E.

KM Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
KM coating composition; glucan; starch; latex; thermoplastic molecule;  
KM amyloplast; vacuole; paper manufacture; mutant; mutein.

OS Streptococcus mutans.  
OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"

XX US2002031826-A1.

XX 14-MAR-2002.

PF 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

PA (NICH/) NICHOLS S E.

XX Nichols SE;

PI WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful  
XX as substitutes for, and additions to modified starch and latexes in paper  
XX manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
XX B polypeptide having changes at position from 148V, D457N, D567T,  
XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,  
XX C 1448V/D457N/D567T/D571K/K79Q/K1014T, Y169N/Y170A/Y171A, and K779Q or a  
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
XX C N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
XX GTF mutant, an isolated polynucleotide, a ribonucleic acid sequence encoding the  
XX complementary polynucleotide, a ribonucleic acid sequence encoding the  
XX GTF mutant, an expression cassette comprising the expression cassette, host  
XX cell introduced with the vector, a transgenic plant, a paper sizing and/or  
XX vector, a seed or tuber from the transgenic plant, a paper transformed  
XX coating composition comprising a glucan produced in a plant transformed  
XX with a gene encoding the mutant GTF, wild type or, starch, a latex,  
XX thermoplastic molecule or their combinations or glucan and starch where  
XX the glucan is produced in the amyloplast and/or vacuole or a maize line  
XX deficient in starch biosynthesis, transformed with a gene encoding a  
XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTFD mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTFD sequence appearing as AAU98029 and the information in  
CC claim 36

XX Sequence 1430 AA;

XX Query Match 76.4%; Score 84; DB 5; Length 1430;

XX Best Local Similarity 85.0%; Pred. No. 1e-05; Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYVFIKRAHSEVOGTRIAKII 22  
||:|||||||  
Db 576 NYVFIKRAHSEVOGTRIAKII 595

RESULT 24  
AAU98041

ID AAU98041 standard; protein; 1430 AA.

AC AAU98041;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFD mutant T589D.

KM Glucosyltransferase; GTFD; transgenic plant; paper sizing;  
KM coating composition; glucan; starch; latex; thermoplastic molecule;  
KM amyloplast; vacuole; paper manufacture; mutant; mutein.

OS Streptococcus mutans.  
OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"

XX US2002031826-A1.

XX 14-MAR-2002.

PF 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

PA (NICH/) NICHOLS S E.

XX Nichols SE;

PI WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.

Claim 36; Page; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D567T, D567T/D571K, D567T/D571K/K1014T, or a 1448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein of potato or the vacuole of sugar beet. is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFD mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFD sequence appearing as AUY98029 and the information in claim 36

Sequence 1430 AA;

Query Match 76.4%; Score 84; DB 5; Length 1430;

Best Local Similarity 85.0%; Pred. No. 1e-05; Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYVFIKRAHSEVQVTRAKII 22

DB 576 NYVFIKRAHSEVQVTRAKII 595

RESULT 25

ID ABR63235 standard; protein; 2055 AA.

XX ABR63235;

DT 23-OCT-2003 (revised)

DT 27-AUG-2003 (first entry)

DE Glucansucrase sequence from strain KG15.

KM Glucan; glucosyltransferase activity; thickener; prebiotic;

XX bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

XX Lactobacillus sakei.

XX WO2003008618-A2.

XX 30-JAN-2003.

XX 22-JUL-2002; 2002WO-NL000495.

XX 20-JUL-2001; 2001EP-00202752.

XX 25-JUL-2001; 2001EP-00202841.

PA (NED) NEDERLANDSE ORG TOEGEPAST.

PI Van Geel - Schutten GH;

DR WPI; 2003-289780/28.

DR N-PSDB; ACC84451.

Novel glucan produced by glucosyltransferase activity of lactic acid bacterium on sucrose substrate, and having backbone consisting of alpha (1,3) - and alpha (1,6) -linked anhydroglucose units, useful as thickener.

Claim 11; Fig 1, 51pp; English.

The present invention relates to glucan capable of being produced by glucosyltransferase activity of a lactic acid bacterium on a sucrose substrate. The method is useful as a thickener, as a prebiotic and as a bioactive agent and as an anti-corrosion agent. The glucan can be incorporated in foodstuffs such as beverages, sauces, dressings, dairy products. The glucan is useful as anticorrosion agent, e.g. for the protection of ship hulls. It can also be incorporated in nutritional or pharmaceutical compositions intended for improving the condition of the gastrointestinal tract. The present sequence represents a sequence of the glucansucrase gene. (Updated on 23-OCT-2003 to standardise OS field)

Sequence 2055 AA;

Query Match 75.5%; Score 83; DB 6; Length 2055;

Best Local Similarity 68.2%; Pred. No. 2.4e-05; Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VENVYFIKRAHSEVQVTRAKII 22

DB 955 VENVYFIKRAHSEVQVTRAKII 976

RESULT 26

ID ABR63229 standard; protein; 221 AA.

XX ABR63229;

DT 27-AUG-2003 (first entry)

DE Glucansucrase sequence from strain Lb86.

KM Glucan; glucosyltransferase activity; thickener; prebiotic;

XX bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

OS Leuconostoc sp.

XX WO2003008618-A2.

XX 30-JAN-2003.

XX 22-JUL-2002; 2002WO-NL000495.

XX 20-JUL-2001; 2001EP-00202752.

XX 25-JUL-2001; 2001EP-00202841.

PA (NED) NEDERLANDSE ORG TOEGEPAST.

PI Van Geel - Schutten GH;

XX WPI; 2003-289780/28.

XX N-PSDB; ACC50073.



vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Sequence 1475 AA;

Query Match 72.7%; Score 80; DB 5; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 5.6e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPNVFIRADSEVQTRIAKII 22  
DB 552 VPSYFIRADSEVQDLIADII 573

RESULT 29  
AAU98040  
ID AAU98040 standard; protein; 1475 AA.

XX AAU98040;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant K779Q.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
XX coating composition; glucan; starch; latex; thermoplastic molecule;  
XX amyloplast; vacuole; paper manufacture; mutant; mutain.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S R.  
XX Nichols SE;  
XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful  
XX as substitutes for and additions to modified starch and latexes in paper  
XX manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44p; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, K148V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Sequence 1475 AA;

Query Match 72.7%; Score 80; DB 5; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 5.6e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPNVFIRADSEVQTRIAKII 22  
DB 552 VPSYFIRADSEVQDLIADII 573

RESULT 30  
AAU98033  
ID AAU98033 standard; protein; 1475 AA.

XX AAU98033;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant K1014T.

KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 XX  
 XX  
 FT Misc-difference 1014  
 FT Location/Qualifiers  
 FT note="Wild-type Lys substituted by Thr"  
 XX  
 XX US2002031826-A1.  
 XX  
 XX  
 PD 14-MAR-2002.  
 PD  
 PD 19-DEC-2000; 2000US-00740274.  
 PD  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009520.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 XX (NICH/) NICHOLS S E.  
 XX  
 XX Nichols SE;  
 XX  
 DR WPI; 2002-414332/44.  
 XX  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 XX Claim 36; Page: 44pp; English.  
 PS  
 PS  
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589S, N471D,  
 CC N471D/T589D, and N471D/T589S. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer

CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 XX Sequence 1475 AA:  
 SQ  
 Query Match 72.7%; Score 80; DB 5; Length 1475;  
 Best Local Similarity 77.3%; Pred. No. 5,6e-05;  
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VENVYFIRAHSEVQTRIAKII 22  
 DB 552 VPSYFIRAHSEVQDPLADII 573  
 Search completed: May 4, 2004, 09:10:04  
 Job time : 50.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:06:31 ; Search time 11.6667 Seconds  
(without alignments)  
181.390 Million cell updates/sec

Title: US-09-290-049a-18  
Perfect score: 110  
Sequence: 1 VPMYVFIKRAHSEVQTRIAKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1365	2 A41483	glucosyltransferase (EC
2	90	81.8	1431	2 A45866	dextranase (EC
3	90	81.8	1508	2 T31098	probable dextran
4	86	78.2	1290	2 JCS473	dextranase (EC
5	80	72.7	1475	2 B31335	glu protein precu
6	77	70.0	1577	2 T30858	glucosyltransferase
7	76	69.1	1375	2 JT0345	glucosyltransferase (EC
8	75	68.2	1599	2 S22737	glucosyltransferase
9	74	67.3	1449	2 T30857	glucosyltransferase
10	74	67.3	1449	2 T30552	glucosyltransferase
11	73	66.4	1518	2 A44811	glucosyltransferase
12	69	62.7	1592	2 A38175	aspartate-tRNA lig
13	49	44.5	602	2 A71256	sugar kinase, FGXY
14	48	43.6	492	2 C72417	aspartate-tRNA lig
15	47	42.7	506	2 E70155	aspartate-tRNA lig
16	47	42.7	557	2 S73434	hypothetical prote
17	45	40.9	498	2 A87374	CBP1 protein - yea
18	45	40.9	654	1 BVBEP1	nine-cis-epoxycar
19	44	40.0	583	2 T04531	conserved hypotet
20	44	40.0	631	2 H90599	probable biotin ca
21	43	39.1	457	2 H71553	acetyl-coenzyme A
22	43	39.1	457	2 D81708	hypothetical prote
23	43	39.1	765	2 S76795	cellulase (EC 3.2.
24	43	39.1	1070	2 S75712	yeast protein homol
25	42.5	38.6	92	2 E47754	conserved hypotet
26	42.5	38.6	319	2 A86777	heat shock protein
27	42	38.2	144	2 H83933	hypothetical prote
28	42	38.2	346	2 E70715	transcription elon
29	41	37.3	160	2 S73763	

30	41	37.3	161	2 B64231	transcription elon
31	41	37.3	245	2 C89986	hypothetical prote
32	41	37.3	270	2 S44952	ImbE protein - Str
33	41	37.3	488	2 B82798	virulence-associat
34	41	37.3	560	1 JIC4795	plasma hyaluronan
35	41	37.3	688	2 T21641	hypothetical prote
36	41	37.3	805	2 T21957	hypothetical prote
37	41	37.3	815	2 T05555	DNA polymerase III
38	41	37.3	934	2 AG1889	WD-40 repeat prote
39	41	37.3	961	1 P1BVA	RNA Ia protein - b
40	41	37.3	1428	2 S62419	hypothetical prote
41	41	37.3	13055	2 T16580	hypothetical prote
42	40	36.4	120	2 H97311	uncharacterized me
43	40	36.4	253	2 A13503	zinc metalloproteid
44	40	36.4	279	2 T27854	hypothetical prote
45	40	36.4	354	2 S65687	(A+T)-stretch-bind
46	40	36.4	594	2 T38114	pyruvate decarboxy
47	40	36.4	595	2 G97179	aspartyl-tRNA synt
48	40	36.4	605	2 T43191	probable pyruvate
49	40	36.4	652	2 T20549	hypothetical prote
50	40	36.4	933	2 G70166	glycoprotein B - I
51	40	36.4	943	2 B48474	hypothetical prote
52	40	36.4	948	2 A56602	probable zinc proc
53	40	36.4	1203	2 T21275	glycoprotein B hom
54	40	36.4	1261	1 A42548	hypothetical prote
55	39.5	35.9	235	2 E86804	genome polypeptid
56	39.5	35.9	289	2 G87673	GntR family transc
57	39.5	35.9	363	2 G83954	hypothetical prote
58	39.5	35.9	366	1 MNXR2	Flagella-associate
59	39.5	35.9	394	2 D95042	sigma NS protein -
60	39.5	35.9	414	2 T06120	UDP-N-acetylglucos
61	39.5	35.9	445	2 B69584	hypothetical prote
62	39	35.5	76	2 G71077	aldehyde dehydroge
63	39	35.5	109	2 JQ1633	HKRF3 protein - hu
64	39	35.5	146	2 T10177	hypothetical prote
65	39	35.5	206	1 B48550	leghemoglobin - ga
66	39	35.5	219	2 E72281	late I3 23k protei
67	39	35.5	236	2 E82593	hypothetical prote
68	39	35.5	252	2 T16934	phosphate regulon
69	39	35.5	264	2 G96978	hypothetical prote
70	39	35.5	285	2 F88281	uncharacterized co
71	39	35.5	288	2 T26383	protein ZK892.1 [I
72	39	35.5	297	2 T28073	hypothetical prote
73	39	35.5	309	2 T49284	hypothetical prote
74	39	35.5	364	2 A96730	hypothetical prote
75	39	35.5	383	2 F69018	formate dehydrogen
76	39	35.5	446	2 AH1957	hypothetical prote
77	39	35.5	488	2 T16402	hypothetical prote
78	39	35.5	495	2 A44489	Gt box-binding pro
79	39	35.5	543	2 G96570	hypothetical prote
80	39	35.5	577	2 D75406	aspartyl-tRNA synt
81	39	35.5	580	1 S33743	aspartate-tRNA lig
82	39	35.5	584	2 D87048	hypothetical prote
83	39	35.5	588	2 B89945	aspartyl-tRNA synt
84	39	35.5	639	2 T31351	adaptor protein CM
85	39	35.5	696	2 T46111	probable transpos
86	39	35.5	879	2 T50388	hypothetical prote
87	39	35.5	1071	2 D86279	hypothetical prote
88	39	35.5	1679	2 T30271	surface protein -
89	39	35.5	2161	1 A45389	genome polypeptid
90	39	35.5	2183	2 S47307	translational activa
91	39	35.5	2672	2 A48126	sigma NS protein -
92	38.5	35.0	366	1 MNXRSD	sigma NS protein -
93	38.5	35.0	366	1 MNXRST	structural protein -
94	38.5	35.0	366	1 C27401	hypothetical prote
95	38.5	35.0	709	1 M2XR3D	hypothetical prote
96	38	34.5	84	2 A83304	hypothetical prote
97	38	34.5	167	2 C64456	hypothetical prote
98	38	34.5	168	2 E83638	hypothetical prote
99	38	34.5	172	2 T34227	hypothetical prote
100	38	34.5	199	2 G75568	conserved hypotet

## ALIGNMENTS

## RESULT 1

A1483

glucosyltransferase (EC 2.4.1.1) gtfS precursor - Streptococcus sobrinus

C/Species: Streptococcus sobrinus

C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C/Accession: A1483

R/Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

Infect. Immun. 58, 2452-2458, 1990

A/Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltrans

A/Reference number: A1483; PMID:9031665; PMID:2142479

A/Accession: A1483

A/Molecule type: DNA

A/Residues: 1-1365 &lt;GTL&gt;

A/Cross-references: GB:M30943; NID:G153652; PIDN:AAA26898.1; PID:G153653

C/Genetics:

A/Status: gtfS

C/Suprafamily: cpl repeat homology

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 100.0%; Score 110; DB 2; Length 1365;

Best Local Similarity 100.0%; Pred. No. 1.2e-09;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNNYFIRAHDSVQTRIAKII 22

DB 537 VNNYFIRAHDSVQTRIAKII 558

## RESULT 2

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C/Species: Streptococcus mutans

C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C/Accession: A45866

R/Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2059-2105, 1990

A/Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A/Reference number: A45866; PMID:91100958; PMID:2148600

A/Accession: A45866

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1431 &lt;HON&gt;

A/Cross-references: GB:M29296

C/Suprafamily: cpl repeat homology

C/Keywords: glycosyltransferase; hexosyltransferase

F/181-201/Domain: cpl repeat homology &lt;CP1&gt;

F/1127-1146/Domain: cpl repeat homology &lt;CP2&gt;

F/1192-1211/Domain: cpl repeat homology &lt;CP3&gt;

F/1257-1276/Domain: cpl repeat homology &lt;CP4&gt;

F/1277-1297/Domain: cpl repeat homology &lt;CP5&gt;

F/1321-1340/Domain: cpl repeat homology &lt;CP6&gt;

F/1341-1361/Domain: cpl repeat homology &lt;CP6&gt;

F/1385-1404/Domain: cpl repeat homology &lt;CP7&gt;

Query Match 81.8%; Score 90; DB 2; Length 1431;

Best Local Similarity 90.0%; Pred. No. 2.5e-06;

Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NNYFIRAHDSVQTRIAKII 22

DB 576 NNYFIRAHDSVQTRIAKII 595

## RESULT 3

probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C/Species: Leuconostoc mesenteroides

C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C/Accession: T31098

R/Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A/Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DE

A/Reference number: Z20981; PMID:98164374; PMID:9503626

A/Accession: T31098

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1508 &lt;MON&gt;

A/Cross-references: EMBL:AF030129; NID:G2766611; PID:G2766612; PIDN:AA895453.1

A/Experimental source: strain NRRL B-1299

C/Genetics:

A/Status: dextr

C/Function: produces dextran composed only of alpha(1-6) glucosidic bonds

C/Description: produces dextran composed only of alpha(1-6) glucosidic bonds

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 81.8%; Score 90; DB 2; Length 1508;

Best Local Similarity 77.3%; Pred. No. 2.7e-06;

Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNNYFIRAHDSVQTRIAKII 22

DB 634 IPNYSFVRAHDSVQTVIAQII 655

## RESULT 4

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C/Species: Leuconostoc mesenteroides

C/Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

C/Accession: J05473

R/Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.

Gene 182, 23-32, 1996

A/Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leucon

A/Reference number: J05473; PMID:97136686; PMID:8982063

A/Accession: J05473

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-1290 &lt;MON&gt;

A/Cross-references: GB:U08181

C/Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont

C/Genetics:

A/Status: dextr

C/Keywords: glycosyltransferase; hexosyltransferase

F/78-870/Domain: catalytic #status predicted &lt;CAT&gt;

F/922-1290/Domain: glucan-binding #status predicted &lt;GB&gt;

Query Match 78.2%; Score 86; DB 2; Length 1290;

Best Local Similarity 85.7%; Pred. No. 1e-05;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 NNYFIRAHDSVQTRIAKII 22

DB 388 NNYFIRAHDSVQTRIAKII 408

## RESULT 5

gtfB protein precursor - Streptococcus mutans

C/Species: Streptococcus mutans

C/Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

C/Accession: B33135

R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A/Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A/Reference number: A33135; PMID:87308013; PMID:3040685

A/Accession: B33135

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1475 &lt;SHI&gt;

A/Cross-references: GB:M17361; NID:G153639; PIDN:AA88588.1; PID:G153640

R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

A/Reference number: A33128

A/Accession: A33128  
 A/Status: preliminary; not compared with conceptual translation  
 A/Description: DNA  
 A/Molecule type: DNA  
 A/Residues: 1-171,173-641,'N',643-1475 <SH2>  
 A/Experimental source: strain GS-5  
 C/Superfamily: cpl repeat homology  
 F/1096-1115/Domain: cpl repeat homology <CP1>  
 F/1224-1243/Domain: cpl repeat homology <CP2>  
 F/1289-1308/Domain: cpl repeat homology <CP3>  
 F/1354-1373/Domain: cpl repeat homology <CP4>  
 F/1419-1438/Domain: cpl repeat homology <CP5>

Query Match 72.7%; Score 80; DB 2; Length 1475;  
 Best Local Similarity 77.3%; Pred. No. 0.00012;  
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPMYVPIRAHDSVQVTRIAKII 22  
 DB 552 VPSYIFIRAHDSVQVTLADI 573

RESULT 6  
 T30858  
 glucosyltransferase - Streptococcus salivarius  
 C/Species: Streptococcus salivarius  
 C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C/Accession: T30858  
 R/Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
 Infect. Immun. 63, 609-621, 1995  
 A/Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pti  
 A/Reference number: Z20909; PMID:95122197; PMID:7822030  
 A/Accession: T30858  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1577 <SIM>  
 A/Cross-references: EMBL:LJ35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
 C/Genetics:  
 A/Gene: gtfm

Query Match 70.0%; Score 77; DB 2; Length 1577;  
 Best Local Similarity 70.0%; Pred. No. 0.00041;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 MYVPIRAHDSVQVTRIAKII 22  
 DB 661 MYIFRAHDSVQVATIANII 680

RESULT 7  
 T30345  
 dextranase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
 N/Alternate names: sucrose 6-glucosyltransferase  
 C/Species: Streptococcus mutans  
 C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
 C/Accession: J0345; C33135  
 R/Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
 Gene 69, 101-109, 1988  
 A/Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
 A/Reference number: J0345; PMID:89137980; PMID:2976010  
 A/Accession: J0345  
 A/Molecule type: DNA  
 A/Residues: 1-1375 <UEN>  
 A/Experimental source: GS-5  
 R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 J. Bacteriol. 169, 4263-4270, 1987  
 A/Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
 A/Reference number: A33135; PMID:87308013; PMID:3040685  
 A/Accession: C33135  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-349 <SHI>  
 A/Cross-references: GB:M17361  
 C/Genetics:

A/Gene: gtfC  
 C/Function: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
 A/Description: cpl repeat homology  
 C/Superfamily: duplication; glycosyltransferase; hexosyltransferase  
 F/1-34/Domain: signal sequence #status predicted <SIG>  
 F/35-1375/Product: glycosyltransferase #status predicted <MAT>  
 F/1126-1145/Domain: cpl repeat homology <CP1>  
 F/1253-1272/Domain: cpl repeat homology <CP2>  
 F/1318-1337/Domain: cpl repeat homology <CP3>

Query Match 69.1%; Score 76; DB 2; Length 1375;  
 Best Local Similarity 72.7%; Pred. No. 0.00051;  
 Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPMYVPIRAHDSVQVTRIAKII 22  
 DB 578 VPSYIFIRAHDSVQVTLIRNII 599

RESULT 8  
 S22737  
 glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
 C/Species: Streptococcus salivarius  
 C/Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
 C/Accession: S22737; S28810; B44811; S22727  
 R/Jacques, N.  
 submitted to the EMBL Data Library, March 1992  
 A/Reference number: S22726  
 A/Accession: S22737  
 A/Molecule type: DNA  
 A/Residues: 1-1599 <JAC>  
 A/Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
 A/Experimental source: ATCC 25975  
 R/Giffard, P.M.; Simpson, C.L.; Millward, C.P.; Jacques, N.A.  
 J. Gen. Microbiol. 137, 2577-2593, 1991  
 A/Title: Molecular characterization of a cluster of at least two glucosyltransferase gen  
 A/Reference number: A44811; PMID:92148377; PMID:1838391  
 A/Accession: S28810  
 A/Molecule type: DNA  
 A/Residues: 1-51 <GIP>  
 A/Cross-references: EMBL:Z11873  
 C/Genetics:  
 A/Gene: gtfK  
 C/Superfamily: cpl repeat homology  
 C/Keywords: glycosyltransferase; hexosyltransferase  
 F/1456-1475/Domain: cpl repeat homology <CPR>

Query Match 68.2%; Score 75; DB 2; Length 1599;  
 Best Local Similarity 78.9%; Pred. No. 0.00088;  
 Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 VYFIRAHDSVQVTRIAKII 22  
 DB 575 YLVPRAHDSVQVTLADI 593

RESULT 9  
 T30857  
 glucosyltransferase - Streptococcus salivarius  
 C/Species: Streptococcus salivarius  
 C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C/Accession: T30857  
 R/Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
 Infect. Immun. 63, 609-621, 1995  
 A/Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pti  
 A/Reference number: Z20909; PMID:95122197; PMID:7822030  
 A/Accession: T30857  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1449 <SIM>  
 A/Cross-references: EMBL:LJ35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
 C/Genetics:



A:Gene: gtfL

Query Match 67.3%; Score 74; DB 2; Length 1449;  
 Best Local Similarity 70.0%; Pred. No. 0.0012;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQTRIAKII 22  
 |||:|||||:|:|:  
 DB 609 NYAFVRAHDSVQSIIGQII 628

RESULT 10

T30552

glucosyltransferase N - Streptococcus salivarius (fragment)  
 C:Species: Streptococcus salivarius  
 C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30552

R:Jaffe, R.I.

submitted to the EMBL Data Library, February 1998

A:Description: Streptococcus salivarius V1477 gtfN.

A:Reference number: Z20854

A:Accession: T30552

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1449 &lt;JAF&gt;

A:Cross-references: EMBL:AF049609; NID:G2935545; PID:G2935546; PIDN:AAC05156.1

C:Genetics:

A:Gene: gtfN

Query Match 67.3%; Score 74; DB 2; Length 1449;  
 Best Local Similarity 70.0%; Pred. No. 0.0012;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQTRIAKII 22  
 |||:|||||:|:|:  
 DB 609 NYAFVRAHDSVQSIIGQII 628

RESULT 11  
 A44811  
 glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

C:Species: Streptococcus salivarius

C&gt;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1993

C:Accession: A44811; S22726; S28809

R:Gifford, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A&gt;Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377; PMID:1838391

A:Accession: A44811

A:Molecule type: DNA

A:Residues: 1-1518 &lt;GIF&gt;

A:Cross-references: EMBL:Z11873; NID:G47526; PIDN:CAA77900.1; PID:G47527

A:Note: Sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)

C:Genetics:

A:Gene: gtfL

C:Superfamily: cpl repeat homology

C:Keywords: glucosyltransferase; hexosyltransferase

F:1307-1326/Domain: cpl repeat homology &lt;CP4&gt;

Query Match 66.4%; Score 73; DB 2; Length 1518;  
 Best Local Similarity 75.0%; Pred. No. 0.0018;  
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQTRIAKII 22  
 |||:|||||:|:|:  
 DB 604 NYVFIKRAHDSVQSIIGQII 623

RESULT 12  
 A38175  
 glucosyltransferase precursor - Streptococcus sobrinus  
 C:Species: Streptococcus sobrinus  
 C>Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999

C:Accession: A38175

R:Ado, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A&gt;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A:Reference number: A38175; MUID:91123227; PMID:1704006

A:Accession: A38175

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1592 &lt;ABO&gt;

A:Cross-references: GB:D90213; NID:G217032; PIDN:BA14241.1; PID:dl014946; PID:G217033

C:Superfamily: cpl repeat homology

F:1083-1112/Domain: cpl repeat homology &lt;CP1&gt;

F:1222-1241/Domain: cpl repeat homology &lt;CP2&gt;

F:1287-1306/Domain: cpl repeat homology &lt;CP3&gt;

F:1330-1351/Domain: cpl repeat homology &lt;CP4&gt;

F:1352-1371/Domain: cpl repeat homology &lt;CP5&gt;

F:1402-1420/Domain: cpl repeat homology &lt;CP6&gt;

F:1465-1484/Domain: cpl repeat homology &lt;CP7&gt;

F:1513-1532/Domain: cpl repeat homology &lt;CP8&gt;

Query Match 62.7%; Score 69; DB 2; Length 1592;  
 Best Local Similarity 68.2%; Pred. No. 0.0087;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPVFIKRAHDSVQTRIAKII 22  
 |||:|||||:|:|:  
 DB 548 VPSYFARAHDSEVDITRDII 569

RESULT 13  
 A71256  
 aspartate-tRNA ligase (EC 6.1.1.12) - syphilis spirochete

N:Alternate names: aspartyl-tRNA synthetase TP0985

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C&gt;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 03-Jun-2002

C:Accession: A71256

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

R.; Ison, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Ueberback, T.; McDo

roney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A&gt;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71256; MUID:98332770; PMID:9665876

A:Accession: A71256

A:Molecule type: DNA

A:Residues: 1-602 &lt;COL&gt;

A:Cross-references: GB:AE001266; GB:AE000520; NID:G3323309; PIDN:AAC65942.1; PID:G332331

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0985

C:Function:

A:Description: activates amino acid and transfers it to specific tRNA molecule

A:Pathway: protein biosynthesis

C:Superfamily: lysine-tRNA ligase

C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 44.5%; Score 49; DB 2; Length 602;  
 Best Local Similarity 56.2%; Pred. No. 5.8; 3; Indels 0; Gaps 0;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 IRANDSVQTRIAKII 22  
 |||:|||||:|:|:  
 DB 505 IRHDTQLOKRIPIIV 520

RESULT 14  
 C72417  
 sugar kinase, FGGV family - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: C72417  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hicke  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratic, M.S.; Phillips, C.A.; Richardson, D.



A:Molecule type: DNA  
A:Residues: 1-654 <VAN>  
A:Cross-references: EMBL:Z49484; NID:G1015590; PIDN:CAA89506.1; PID:G1015591; GSPDB:GN00  
R:Purnelle, B.; Coster, F.; Goffeau, A.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56977  
A:Accession: S56996  
A:Molecule type: DNA  
A:Residues: 637-654 <PUR>  
A:Cross-references: EMBL:Z49484; GSPDB:GN00010; MIPS:YJL209W  
C:Genetics:  
A:Gene: SGD:CBP1, MIPS:YJL209W  
A:Cross-references: SGD:S0003745, MIPS:YJL209W  
A:Map position: 10L  
A:Genome: nuclear  
C:Function:  
A:Description: pre-mRNA processing  
A>Note: required for correct 5' terminal processing of cytochrome b pre-mRNA  
C:Superfamily: CBP1 protein  
C:Keywords: mitochondrion

Query Match 40.9%; Score 45; DB 1; Length 654;  
Best Local Similarity 47.4%; Pred. No. 29;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQTRIAKI 21  
: ||| : : : :  
Db 300 HYVISAHSHRIYKVKFI 318

RESULT 19  
T04531  
nine-cis-epoxycarotenoid dioxygenase homolog F28J12.10 - Arabidopsis thaliana  
N/Alternate names: hypothetical protein T9A21.200  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
C/Accession: T04531; T04937  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duescherhoef, A.; Bancroft,  
submitted to the Protein Sequence Database, February 1998  
A:Reference number: Z15377  
A:Accession: T04531  
A:Molecule type: DNA  
A:Residues: 1-583 <BEV>  
A:Cross-references: EMBL:AL021710  
A:Experimental source: cultivar Columbia; BAC clone F28J12  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15390  
A:Accession: T04937  
A:Molecule type: DNA  
A:Residues: 1-377 <BRW>  
A:Cross-references: EMBL:AL021713  
A:Experimental source: cultivar Columbia; BAC clone T9A21  
C:Genetics:  
A:Map position: 4  
A>Note: F28J12.10; T9A21.200

Query Match 40.0%; Score 44; DB 2; Length 583;  
Best Local Similarity 42.1%; Pred. No. 38;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 4 YVFIRAHDSVQTRIAKI 22  
: ||| : : : :  
Db 530 YIFCHVDETKTSELQII 548

RESULT 20  
H90599  
conserved hypothetical protein MYPU\_7040 [imported] - Mycoplasma pulmonis (strain UAB CT  
C/Species: Mycoplasma pulmonis  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: H90599  
R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001  
A>Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: H90599  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-631 <KUR>  
A:Cross-references: GB:AL445566; PID:G14090119; PIDN:CAIC13877.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPU\_7040  
A:Genetic code: SGC3  
C:Superfamily: conserved hypothetical protein MG139

Query Match 40.0%; Score 44; DB 2; Length 631;  
Best Local Similarity 26.7%; Pred. No. 41;  
Matches 4; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPNTVFIRAHDSVQ 15  
: ||| : : : :  
Db 49 IPNYSYLKHEHQIE 63

RESULT 21  
H71553  
probable biotin carboxylase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C/Species: Chlamydia trachomatis  
C/Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Jan-2000  
C/Accession: H71553  
R:Stephens, R.S.; Kalman, S.; Lammel, C.U.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
Science 282, 754-759, 1998  
A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac  
A:Reference number: AY1570; MUID:9900809; PMID:9784136  
A:Accession: H71553  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-457 <ARN>  
A:Cross-references: GB:AE001286; GB:AE001273; NID:G3328516; PIDN:AAAC67715.1; PID:G33285;  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: accC  
C:Superfamily: biotin carboxylase; biotin carboxylase homology  
F.3-448/Domain: biotin carboxylase homology <ECH>

Query Match 39.1%; Score 43; DB 2; Length 457;  
Best Local Similarity 45.0%; Pred. No. 42;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQTRIAKI 22  
: ||| : : : :  
Db 219 NYVYIGERDCTVQRRQKLI 238

RESULT 22  
DB1708  
acetyl-coenzyme A carboxylase, biotin carboxylase TC0400 [imported] - Chlamydia muridar  
C/Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: DB1708  
R:Read, T.D.; Brummam, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: AB1500; MUID:20150255; PMID:10684935  
A:Accession: DB1708  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-457 <TET>  
A:Cross-references: GB:AE002306; GB:AE002160; NID:G7190428; PIDN:AAF39257.1; PID:G71904  
A:Experimental source: strain Ni9g (MoPn)  
C:Genetics:  
A:Gene: TC0400  
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match 39.1%; Score 43; DB 2; Length 457;  
Best Local Similarity 45.0%; Pred. No. 42;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQTRIAKI 22  
DB 219 NYVYIGERDCTVQRRKRI 238

## RESULT 23

S76795  
Hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variate: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S76795  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76795  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-765 <KAN>

A:Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDD:BA18707.1; PID:d101944  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 39.1%; Score 43; DB 2; Length 765;  
Best Local Similarity 53.3%; Pred. No. 76;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 TRAHDSVQTRIAKI 21  
DB 463 VRALQTRIQRIANL 477

## RESULT 24

S75712  
Cellulase (EC 3.2.1.4) - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: endo-1,4-beta-glucanase; protein slr0897  
C:Species: *Synechocystis* sp.

A:Variate: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S75712  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S75712  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-1070 <KAN>  
A:Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDD:BA10447.1; PID:d101105  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Species: *Synechocystis* sp.  
A:Superfamily: bacterial cellulose-binding domain homology  
C:Keywords: glycosidase; hydrolase  
F:7-112/Domain: bacterial cellulose-binding domain homology <BCB2>  
F:434-539/Domain: bacterial cellulose-binding domain homology <BCB2>

Query Match 39.1%; Score 43; DB 2; Length 1070;  
Best Local Similarity 50.0%; Pred. No. 11e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 NYVFIKRAHDSVQTRIA 18  
DB 244 NYVYIRSHYGVQERI 259

## RESULT 25

E47754  
Ycea protein homolog (3' to hlsE) - *Lactococcus lactis* subsp. *lactis* (fragment)

C:Species: *Lactococcus lactis* subsp. *lactis*  
C:Date: 24-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 13-Nov-1998

C:Accession: E47754  
R:Delorme, C.; Ehrlich, S.D.; Renault, P.  
J. Bacteriol. 174, 6571-6579, 1992  
A:Title: Histidine biosynthesis genes in *Lactococcus lactis* subsp. *lactis*.

A:Reference number: A45734; MUID:93015709; PMID:1400209

A:Contents: NC002118

A:Accession: E47754  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-92 <DEL>

A:Note: sequence extracted from NCBI backbone (NCBIN:115814, NCBI:P:115837)  
C:Superfamily: *Bacterichia coli ycea* protein

Query Match 38.6%; Score 42.5; DB 2; Length 92;  
Best Local Similarity 47.6%; Pred. No. 7.8;  
Matches 10; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

QY 2 PNRYFIKRAHDSVQTRIA 19  
DB 64 PNRYFIKRAHDSVQTRIA 84

## RESULT 26

A86777  
conserved hypothetical protein ymdE [imported] - *Lactococcus lactis* subsp. *lactis* (strain

C:Species: *Lactococcus lactis* subsp. *lactis*  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: A86777  
R:Belcic, A.; Winkler, P.; Mager, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich,

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp

A:Reference number: A86625; MUID:2125186; PMID:11337471

A:Accession: A86777  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-319 <STO>

A:Cross-references: GB:AE005176; PID:g12724186; PIDD:AAK05315.1; GSPDB:GN00146

A:Experimental source: strain IL1403  
C:Genetics:

A:Superfamily: *Bacterichia coli ycea* protein

Query Match 38.6%; Score 42.5; DB 2; Length 319;  
Best Local Similarity 47.6%; Pred. No. 33;  
Matches 10; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

QY 2 PNRYFIKRAHDSVQTRIA 19  
DB 291 PNRYFIKRAHDSVQTRIA 311

## RESULT 27

H83933  
heat shock protein class I (low molecular weight) BH2272 [imported] - *Bacillus halodurans*

C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: H83933  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: H83933  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-144 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDD:BA005991.1; GSPDB:GN00

A:Experimental source: strain C-125  
C:Genetics:

A:Gene: BH2272

Query Match 38.2%; Score 42; DB 2; Length 144;  
 Best Local Similarity 61.5%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PNVPFIRAHSEVOTRIAKI 14  
 DB 74 PÖYVRIRIHREEV 86

RESULT 28

E70715  
 hypothetical protein RV0943c - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: E70715

R:Colo, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Rajandream, M.A.; Rogers, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Nature 353, 537-544, 1998  
 A:Authors: Sgares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; PMID:98255987; PMID:9634230  
 A:Accession: E70715

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-346 <COL>

A:Cross-references: GB:Z79700; GB:AL123456; NID:g3261628; PIDN:CAB01978.1; PID:e264106;  
 A:Experimental source: strain H37Rv

C:Genetics:  
 A:Gene: RV0943c

Query Match 38.2%; Score 42; DB 2; Length 346;  
 Best Local Similarity 50.0%; Pred. No. 44;  
 Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 PNVPFIRAHSEVOTRIAKI 17  
 DB 177 PNVPFIRAHSEVOTRIAKI 192

RESULT 29

S73763  
 transcription elongation factor greA - Mycoplasma pneumoniae (strain ATCC 29342)  
 N:Alternate names: hypothetical protein F11\_orf160  
 C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342  
 C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 C:Accession: S73763  
 R:Himmelfreid, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S73727; PMID:97105885; PMID:8948633  
 A:Accession: S73763

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-160 <HIM>

A:Cross-references: EMBL:AE000043; GB:U00089; NID:g1674122; PIDN:AA86085.1; PID:g167412  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:

A:Gene: greA  
 A:Genetic code: SGC3  
 C:Superfamily: transcription elongation factor greB  
 C:Keywords: transcription factor

Query Match 37.3%; Score 41; DB 2; Length 160;  
 Best Local Similarity 42.1%; Pred. No. 26;  
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIRAHSEVOTRIAKI 21  
 DB 49 DYDAKAKOGGRIETRIAKI 67

RESULT 30

B64231  
 transcription elongation factor greA homolog - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: B64231

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhmann, J.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.; C.A.; Venter, J.C.  
 Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; PMID:96026346; PMID:7569993

A:Accession: B64231  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-161 <TRI>  
 A:Cross-references: GB:U039708; GB:LA3967; NID:g3844869; PIDN:AACT1504.1; PID:g1045978; T

C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: transcription elongation factor greB  
 C:Keywords: transcription factor

Query Match 37.3%; Score 41; DB 2; Length 161;  
 Best Local Similarity 42.1%; Pred. No. 26;  
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIRAHSEVOTRIAKI 21  
 DB 49 DYDAKAKOGGRIETRIAKI 67

Search completed: May 4, 2004, 09:13:22  
 Job time: 14.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: May 4, 2004, 09:02:41; Search time 8 Seconds

(Without alignments)  
143.193 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VPNNVFIRAHDSVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1365	1	GTFP5_STRDO
2	90	81.8	1462	1	GTFP2_STRDO
3	76	69.1	1455	1	GTFP3_STRDO
4	75	68.2	1476	1	GTFP4_STRDO
5	70	63.6	1597	1	GTFP1_STRDO
6	69	62.7	1592	1	GTFP2_STRDO
7	49	44.5	602	1	SYD_TREPA
8	47	42.7	336	1	RAFI_DROKE
9	47	42.7	557	1	SYD_MYCPN
10	47	42.7	586	1	SYD_BORBU
11	47	42.7	592	1	SYD_TRETN
12	45	40.9	654	1	CBPI_YEAST
13	43	39.1	406	1	AROA_YEAST
14	43	39.1	431	1	ARGD_BIFLO
15	42.5	38.6	319	1	IMDE_LACIA
16	42	38.2	346	1	Y943_MYCTU
17	41.5	37.7	353	1	DPO4_SULTO
18	41	37.3	160	1	GRE4_MYCPN
19	41	37.3	161	1	GRE4_MYCPN
20	41	37.3	223	1	KAD4_MOUSE
21	41	37.3	223	1	KAD4_MOUSE
22	41	37.3	339	1	Y538_STRPY
23	41	37.3	584	1	SYD_BUCAP
24	41	37.3	630	1	COG6_DROWE
25	41	37.3	737	1	SKNI_CANAL
26	41	37.3	961	1	VIA_BMV
27	41	37.3	1428	1	Y484_SCHPO
28	40	36.4	146	1	V224_FOWPV
29	40	36.4	205	1	ADEN_ADEG8
30	40	36.4	580	1	SYD_MYCPN
31	40	36.4	594	1	DCP2_SCHPO
32	40	36.4	595	1	SYD_GLOBB
33	40	36.4	973	1	RRPO_PAV

34	40	36.4	1553	1	OSP_DROME	Q27421 drosophila
35	40	36.4	2261	1	RRPL_MUMPM	P30929 mumps virus
36	39.5	35.9	366	1	RRPO_REOVL	P12002 reovirus (t
37	39.5	35.9	445	1	DHA3_BACSV	P46329 bacillus su
38	39	35.5	145	1	LG85_PEA	O48655 pistum activ
39	39	35.5	206	1	ADEN_ADECU	P35990 canine aden
40	39	35.5	236	1	PHOU_XYTRA	Q95191 xylella fas
41	39	35.5	297	1	LECG_CABEL	Q09581 caenorhabdi
42	39	35.5	577	1	SYD_DEIRA	Q91071 deinococcus
43	39	35.5	580	1	SYD_TRETN	P36419 thermus the
44	39	35.5	588	1	SYD_STRAM	Q99119 staphylococ
45	39	35.5	606	1	SP2_HUMAN	O86899 staphylococ
46	39	35.5	637	1	CZAP_MOUSE	Q02086 homo sapien
47	39	35.5	637	1	CZAP_MOUSE	Q91190 mus musculu
48	39	35.5	751	1	EFGL_HUMAN	Q95K65 homo sapien
49	39	35.5	879	1	PMCI_SCHPO	Q997Y4 schizosacch
50	39	35.5	2183	1	RRPL_RINDR	P41357 rinderpest
51	39	35.5	2672	1	RRPL_CDVO	P24658 canine dist
52	39	35.5	2672	1	RRPL_CDVO	P33892 saccharomyc
53	39	35.5	366	1	RRPO_REOVL	P07940 reovirus (t
54	38.5	35.0	366	1	RRPO_REOVL	P1078 reovirus (t
55	38.5	35.0	709	1	YCS1_METJA	Q925Y7 rhizobium m
56	38.5	35.0	167	1	YCS1_METJA	O8m011 coxynebacte
57	38	34.5	217	1	MSA1_RHIME	Q914T8 halobacteri
58	38	34.5	251	1	LTBP_CORGL	P73348 synchocyst
59	38	34.5	251	1	LTBP_CORGL	O98c24 shewanella
60	38	34.5	267	1	IF2A_HALNI	P74548 synchocyst
61	38	34.5	289	1	DHPS_SYNY3	Q895X7 buchnera ap
62	38	34.5	336	1	DDL_SHEON	P49682 homo sapien
63	38	34.5	346	1	SYPA_LACIA	Q92121 rickettsia
64	38	34.5	355	1	CYSA_SYNY3	O05966 rickettsia
65	38	34.5	360	1	HIS6_BUCBP	O89Z93 yersinia pe
66	38	34.5	368	1	CCR3_HUMAN	P47493 mycoplasma
67	38	34.5	386	1	SUCC_RICCN	Q839V7 enterococcu
68	38	34.5	386	1	SUCC_RICCN	O67820 aquifex ae
69	38	34.5	401	1	PCMB_YERPE	O67820 aquifex ae
70	38	34.5	446	1	SYG_MYCPN	O9c6f8 lactococcu
71	38	34.5	485	1	SYE_ENTRA	O7muf7 porphyromon
72	38	34.5	490	1	IMDH_AQUNE	O99Kc1 bacillus ha
73	38	34.5	501	1	IMDH_AQUNE	Q91558 pseudomonas
74	38	34.5	505	1	SYE_FORGI	P53037 saccharomyc
75	38	34.5	637	1	MUTL_BACDI	P11223 avian infec
76	38	34.5	641	1	PRIM_UREPA	P12650 avian infec
77	38	34.5	651	1	ACSI_PSEBE	P47767 spiroplasma
78	38	34.5	1138	1	PSD2_YEAST	P52179 homo sapien
79	38	34.5	1162	1	VG12_IBVK	Q61026 mus musculu
80	38	34.5	1162	1	VG12_IBVK	O15596 homo sapien
81	38	34.5	1162	1	VG12_IBVK	O62234 mus musculu
82	38	34.5	1302	1	PROB_SPTCI	O00799 plasmodium
83	38	34.5	1451	1	MYM1_HUMAN	Q15149 homo sapien
84	38	34.5	1462	1	NC02_HUMAN	P16369 drosophila
85	38	34.5	1462	1	NC02_HUMAN	P42888 synechobact
86	38	34.5	1666	1	RRP2_PLAVB	O27092 methanobact
87	38	34.5	2867	1	RRP2_PLAVB	O40545 nicotiana t
88	38	34.5	4499	1	DYHA_CHLRE	P49361 flavetia pr
89	38	34.5	4684	1	PLEI_HUMAN	O49852 flavetia an
90	37.5	34.1	192	1	CUPP_DROPS	P49361 flavetia pr
91	37.5	34.1	344	1	CYSA_SYNY7	P61901 porcine rot
92	37.5	34.1	372	1	PR52_MERTH	Q92799 chlamydia p
93	37.5	34.1	593	1	PR52_MERTH	
94	37.5	34.1	593	1	PR52_MERTH	
95	37.5	34.1	1034	1	GCSP_FLAPR	
96	37.5	34.1	1034	1	GCSP_FLAPR	
97	37.5	34.1	1034	1	GCSP_FLAPR	
98	37.5	34.1	1037	1	GCSP_FLAPR	
99	37.5	34.1	1082	1	RRPO_ROTPC	
100	37.5	34.1	111	1	RL23_CHLPP	

## ALIGNMENTS

RESULT 1

ID	GENES	STANDARD	PERT	1365 AA.
AC	P29335			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)			
DN	(Sucrose 6-glucosyltransferase).			
OS	Streptococcus downei (Streptococcus sobrinus).			
OC	Bacteriae; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_Taxid=1317;			
RY	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ME29;			
RX	MEDLINE=90316665; PubMed=2142479;			
RA	Glimore R.S., Ruseell R.R., Ferret J.J.;			
RT	"Analysis of the Streptococcus downei gtfS gene, which specifies a			
RL	glucosyltransferase that synthesizes soluble glucans".			
RL	Infect. Immun. 58:2452-2458(1990).			
CC	-1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT			
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE			
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE			
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	-1- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl} (N-1)			
CC	fructose + {(1,6)-alpha-D-glucosyl} (N-1)			
CC	-1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF			
CC	PRIMER GLUCAN UNLIKE GTF-1.			
CC	-1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA			
CC	1,6-GLUCOSE).			
CC	-1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.			
CC	-1- SIMILARITY: Contains 10 cell wall binding repeats.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; M30943; AAA26898.1; -			
DR	InterPro; IPR0023479; CW_binding.			
DR	InterPro; IPR003318; Glyco_hydro_70.			
DR	Pfam; PF01473; CW_binding_1; 8.			
DR	Pfam; PF02324; Glyco_hydro_70; 1.			
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.			
FT	SIGNAL	1	36	OR 37 (POTENTIAL).
FT	CHAIN	37	1365	GLUCOSYLTRANSFERASE-S.
FT	REPEAT	157	177	CELL WALL BINDING 1.
FT	REPEAT	178	197	CELL WALL BINDING 2.
FT	DOMAIN	198	1061	CATALYTIC (APPROXIMATE).
FT	REPEAT	1062	1082	CELL WALL BINDING 3.
FT	REPEAT	1083	1102	CELL WALL BINDING 4.
FT	REPEAT	1150	1169	CELL WALL BINDING 5.
FT	REPEAT	1170	1190	CELL WALL BINDING 6.
FT	REPEAT	1225	1243	CELL WALL BINDING 7.
FT	REPEAT	1289	1308	CELL WALL BINDING 8.
FT	REPEAT	1309	1328	CELL WALL BINDING 9.
FT	REPEAT	1331	1352	CELL WALL BINDING 10.
SO	SEQUENCE	1365 AA;	151590 MW;	16729685A2E8C476 CRC64;
Query Match		100.0%;	Score 110;	DB 1; Length 1365;
Best Local Similarity		100.0%;	Pred. No. 6; 9e-10;	
Matches	22;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	VPNVFIKRAHDSVQIRIAKII	22	
Db	537	VPNVFIKRAHDSVQIRIAKII	558	

```

GTFD_STEMU STANDARD; PRT; 1462 AA.
ID _GTFD_STEMU STANDARD; PRT; 1462 AA.
AC P49331, O693983, O69386, O69389, O69392, O69396;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
DE GTFD OR SMU 910.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1309;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RC MEDLINE=91100958; PubMed=2148600;
RA Honda O., Kato C., Kuramitsu H.K.;
RA "Nucleotide sequence of the Streptococcus mutans gtf gene encoding
RT the glucosyltransferase-S enzyme."
RT J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,
RC MT4467 / Serotype B, and MT8148 / Serotype C;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kikuma S., Hamada S.;
RA "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans."
RT FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RA "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RL
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDiate THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl} (N) = D-
CC fructose + {(1,6)-alpha-D-glucosyl} (N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 6 cell wall binding repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```



DR PFam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries;  
 FT Complete proteome.  
 FT SIGNAL 1  
 FT CHAIN ? 1462  
 FT DOMAIN 1332 1423  
 FT REPEAT 1232 1295  
 FT REPEAT 1296 1359  
 FT REPEAT 1360 1423  
 FT VARIANT 10 10  
 FT VARIANT 19 19  
 FT VARIANT 58 58  
 FT VARIANT 68 68  
 FT VARIANT 81 81  
 FT VARIANT 113 113  
 FT VARIANT 122 122  
 FT VARIANT 132 132  
 FT VARIANT 135 135  
 FT VARIANT 137 137  
 FT VARIANT 202 202  
 FT VARIANT 255 255  
 FT VARIANT 275 275  
 FT VARIANT 288 288  
 FT VARIANT 301 301  
 FT VARIANT 313 313  
 FT VARIANT 317 317  
 FT VARIANT 328 328  
 FT VARIANT 350 350  
 FT VARIANT 628 628  
 FT VARIANT 688 688  
 FT VARIANT 726 726  
 FT VARIANT 730 730  
 FT VARIANT 762 762  
 FT VARIANT 964 964  
 FT VARIANT 1019 1019  
 FT VARIANT 1059 1060  
 FT VARIANT 1060 1060  
 FT VARIANT 1080 1080  
 FT VARIANT 1142 1142  
 FT VARIANT 1198 1198  
 FT VARIANT 1220 1220  
 FT VARIANT 1280 1280  
 FT VARIANT 1282 1282  
 FT VARIANT 1290 1290  
 FT VARIANT 1311 1311  
 FT VARIANT 1403 1403  
 FT VARIANT 1425 1425  
 FT VARIANT 1449 1449  
 FT VARIANT 1462 1462  
 FT CONFLICT 1462 1462  
 FT SEQUENCE 1462 AA; 163387 MW; CE4A279C4D708645 CRC64;  
 SQ R (IN REF. 1).  
 Query Match 81.8%; Score 90; DB 1; Length 1462;  
 Best Local Similarity 90.0%; Pred. No. 14e-06;  
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 3 NYVFIRADSEVOTRIAKII 22  
 Db 576 NYVFIRADSEVOTRIAKII 595  
 RESULT 3  
 GTF\_C\_STRMU STANDARD; PRT; 1455 AA.

AC P13470; 069382; 069385; 069388; 069391; 069397; P05427;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)  
 DE (dextranucrase) (sucrose 6-glucosyltransferase).  
 GN GTF\_C OR SMU.1005.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RX MEDLINE=89137980; PubMed=2976010;  
 RT Ueda S., Shiroza T., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5";  
 RT Gene 69:101-109(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,  
 RC MT4467 / Serotype B, and MT8148 / Serotype C;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans";  
 RT FEMS Microbiol. Lett. 161:331-336(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
 RX MEDLINE=22295063; PubMed=12397186;  
 RA Ajdic D., McShan W.M., McLaughlin R.B., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Ehn S., Qian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Perlecci J.U.;  
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
 RT pathogen";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 RN [4]  
 RP SEQUENCE OF 1-349 FROM N.A.  
 RC STRAIN=GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans";  
 RT J. Bacteriol. 169:4263-4270(1987).  
 RN [5]  
 RP FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 RP TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 RP OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE  
 RP AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6)-alpha-D-glucosyl(N) = D-  
 CC fructose + (1,6)-alpha-D-glucosyl(N+1).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: GTF-SI SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-SI SYNTHESIZES  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
 CC -1- SIMILARITY: Contains 5 cell wall binding repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; M22054; AAA88592.1; -  
 CC EMBL; D88652; BAA26102.1; -  
 CC EMBL; D88655; BAA26106.1; -  
 CC EMBL; D88658; BAA26110.1; -  
 CC EMBL; D88661; BAA26114.1; -  
 CC EMBL; D88978; BAA26120.1; -



```

DR EMBL: AE014940; AAN58706.1; -.
DR EMBL: M17361; AAB8589.1; -.
DR PIR: J070345; J070345.
DR InterPro: IPR002479; CM binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CM binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries;
KW Complete proteome.
FT SIGNAL 1 34
FT CHAIN 35 1455
FT DOMAIN 35 1050
FT 1126 1455
FT 1126 1455
FT DOMAIN 1126 1455
FT 1126 1455
FT REPEAT 1169 1200
FT REPEAT 1227 1238
FT REPEAT 1253 1303
FT REPEAT 1318 1330
FT 21 21
FT 81 81
FT 106 106
FT 116 116
FT 126 126
FT 150 151
FT 256 256
FT 425 425
FT 519 519
FT 538 538
FT 545 545
FT 597 597
FT 600 600
FT 601 601
FT 614 614
FT 727 727
FT 734 734
FT 964 964
FT 1113 1113
FT 1118 1118
FT 1204 1204
FT 1208 1208
FT 1292 1294
FT 1305 1369
FT 1326 1326
FT 1331 1331
FT 1377 1377
FT 1398 1398
FT 1424 1424
FT 1439 1439
FT 1444 1444
FT 1455 1455
FT 1455 AA; 162965 MM; 3CB455A9A4FEC86 CEC64;
SEQUENCE 1455 AA; 162965 MM; 3CB455A9A4FEC86 CEC64;
Query Match 69.1%; Score 76; DB 1; Length 1455;
Best Local Similarity 72.7%; Pred. No. 0.00028;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 VENVYPIRAHDSVQIRIAKII 22
Db 578 VPSYSPIRAHDSVQDLIRNII 599

```

```

GTFB_STRMU STANDARD; PRT: 1476 AA.
ID GTFB_STRMU 069381; 069384; 069390; 069396;
AC P08987; 069381; 069384; 069390; 069396;
DT 01-NOV-1998 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
DE GTFB OR SMU 1004.
OS Streptococcus mutans.
CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CC NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87306013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT Shiroza T., Ueda S., Kuramitsu H.K.;
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,
RX MT4467 / Serotype B, and MT8148 / Serotype C;
RA MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans."
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=2295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Jin S., Qian Y.,
RA Li S., Zhu H., Nejauf F., Lai H., White J., Roe B.A., Ferretti J.U.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6)-alpha-D-glucosyl (N) = D-
CC fructose + (1,6)-alpha-D-glucosyl (N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 10 cell wall binding repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M17361; AAB8589.1; -.
DR EMBL: D88651; BAA26109.1; -.
DR EMBL: D88654; BAA26105.1; -.
DR EMBL: D88657; BAA26109.1; -.
DR EMBL: D88660; BAA26113.1; -.
DR EMBL: D89977; BAA26119.1; -.
DR EMBL: AE014940; AAN58705.1; -.
DR PIR: B31335; B31335.
DR InterPro: IPR002479; CM binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CM_binding_1; 7.

```

DR Pfam; PF02324; Glyco\_hydro 70; 1.  
 KM Transferase; Glycosyltransferase; Signal; Repeat; Dental caries;  
 KW Complete proteome.

FT	SIGNAL	1	34	POTENTIAL.
FT	CHAIN	35	1476	GLUCOSYLTRANSFERASE-1.
FT	DOMAIN	35	1051	CATALYTIC (APPROXIMATE).
FT	DOMAIN	1097	1476	GLUCAN-BINDING (APPROXIMATE).
FT	REPEAT	1097	1130	A REPEAT
FT	DOMAIN	1161	1470	5 X TANDEM REPEATS.
FT	REPEAT	1161	1210	1.
FT	REPEAT	1225	1275	2.
FT	REPEAT	1290	1340	3.
FT	REPEAT	1355	1405	4.
FT	REPEAT	1420	1470	5.
FT	VARIANT	62	62	S -> T (IN STRAIN MT4239).
FT	VARIANT	65	65	T -> I (IN STRAIN GS-5).
FT	VARIANT	68	68	V -> A (IN STRAINS GS-5, MT4245, MT4251, MT4467 AND MT8148).
FT	VARIANT	78	78	O -> P (IN STRAIN MT4251).
FT	VARIANT	86	86	I -> S (IN STRAINS GS-5, MT4245, MT4251, MT4467 AND MT8148).
FT	VARIANT	89	89	S -> F (IN STRAIN MT4251).
FT	VARIANT	168	168	K -> N (IN STRAIN MT4251).
FT	VARIANT	276	276	S -> D (IN STRAINS GS-5, MT4467 AND MT8148).
FT	VARIANT	399	399	N -> R (IN STRAIN MT4239).
FT	VARIANT	474	474	I -> T (IN STRAIN MT4239).
FT	VARIANT	512	512	K -> R (IN STRAIN MT8148).
FT	VARIANT	519	519	F -> I (IN STRAIN MT8148).
FT	VARIANT	701	701	T -> I (IN STRAIN MT8148).
FT	VARIANT	708	708	A -> V (IN STRAIN MT8148).
FT	VARIANT	938	938	P -> L (IN STRAIN MT8148).
FT	VARIANT	952	957	FGKVE -> YGTPVA (IN STRAINS GS-5, MT4239 AND MT4467).
FT	VARIANT	963	964	SV -> NT (IN STRAINS GS-5, MT4239 AND MT4467).
FT	VARIANT	968	970	ADS -> VDG (IN STRAINS GS-5, MT4239 AND MT4467).
FT	VARIANT	1086	1086	A -> T (IN STRAIN MT4239).
FT	VARIANT	1158	1158	S -> N (IN STRAIN MT4239).
FT	VARIANT	1163	1163	H -> Y (IN STRAIN MT4251).
FT	VARIANT	1168	1168	E -> K (IN STRAIN MT8148).
FT	VARIANT	1182	1182	Y -> C (IN STRAIN MT8148).
FT	VARIANT	1234	1234	A -> P (IN STRAIN MT4239).
FT	VARIANT	1263	1263	R -> P (IN STRAIN MT8148).
FT	VARIANT	1263	1263	R -> H (IN STRAINS GS-5 AND MT4467).
FT	VARIANT	1264	1264	Y -> H (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).
FT	VARIANT	1272	1272	S -> G (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).
FT	VARIANT	1329	1329	H -> Y (IN STRAINS GS-5 AND MT4467).
FT	VARIANT	1394	1394	Y -> H (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).
FT	VARIANT	1402	1402	S -> G (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).
FT	VARIANT	1459	1459	Y -> H (IN STRAIN MT4467).
FT	CONFLICT	570	570	R -> A (IN REF. 1).
FT	CONFLICT	800	817	ADQDVAASTAPSTDGK -> LIRFALRLRPHQOMA (IN REF. 1).
FT	CONFLICT	1310	1310	H -> L (IN REF. 1).
FT	CONFLICT	1476	1476	H -> L (IN REF. 1).
FT	SEQUENCE	1476	1476	AA; 165846 MW; 9C6E09F31B4C6C CRC64;

Query Match 68.2%; Score 75; DB 1; Length 1476;  
 Best Local Similarity 72.7%; Pred. No. 0.0042;  
 Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 VPNVYFIRAHDSVQTRIAKII 22  
 ||:|||||||  
 DB 552 VPSYFIRAHDSVQDILRDII 573

ID GTF1\_STRDO STANDARD; PRT; 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucosyltransferase-1 precursor (EC 2.4.1.5) (GTF-1) (Dextranucrase)  
 DE (Sucrose 6-glucosyltransferase).  
 GN GTF1.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_Taxid=1317;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 RT sobrinus MFE28.";  
 RL J. Bacteriol. 169:4271-4278 (1987).  
 CC  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl] (N) = D-  
 CC fructose + [(1,6)-alpha-D-glucosyl] (N+1).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
 CC -1- SIMILARITY: Contains 19 cell wall binding repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC  
 CC EMBL; M17391; AAC63063.1; -;  
 CC InterPro; IPR002479; CW binding.  
 CC DR Pfam; PF01473; CW binding 1; 16.  
 CC DR Pfam; PF02324; Glyco\_hydro 70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT	SIGNAL	1	38	POTENTIAL.
FT	CHAIN	39	1597	GLUCOSYLTRANSFERASE-1.
FT	DOMAIN	39	1050	CATALYTIC (APPROXIMATE).
FT	DOMAIN	1099	1597	GLUCAN-BINDING (APPROXIMATE).
FT	REPEAT	1099	1132	1-25 A, 2 B, AND 5 AC REPEATS.
FT	REPEAT	1163	1213	A REPEAT.
FT	REPEAT	1227	1277	AC REPEAT.
FT	REPEAT	1292	1342	AC REPEAT.
FT	REPEAT	1352	1399	B REPEAT.
FT	REPEAT	1406	1455	AC REPEAT.
FT	REPEAT	1465	1512	B REPEAT.
FT	REPEAT	1519	1568	AC REPEAT.
FT	REPEAT	1582	1597	A REPEAT (INCOMPLETE).
FT	SEQUENCE	1597	1597	AA; 177080 MW; B9E6A200868798R CRC64;

Query Match 63.6%; Score 70; DB 1; Length 1597;  
 Best Local Similarity 68.2%; Pred. No. 0.003;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 VPNVYFIRAHDSVQTRIAKII 22  
 ||:|||||||  
 DB 554 VPSYFIRAHDSVQDILRDII 575

RESULT 6  
GTF2\_STRDO STANDARD; PRT; 1592 AA.  
ID GTF2\_STRDO  
AC P27470;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
OS (Streptococcus downei (Streptococcus downei)).  
DE Streptococcus downei (Streptococcus downei).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus;  
CX NCBI\_TaxID=1317;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6715 / Serotype G;  
RX MEDLINE=91123227; PubMed=1704006;  
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
RA Kagawa H.;  
RT "Peptide sequences for sucrose splitting and glucan binding within  
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
RT synthetase).";  
RL J. Bacteriol. 173:989-996(1991).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC OF PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl] (N) = D-  
CC fructose + [(1,6)-alpha-D-glucosyl] (N+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
CC -1- SIMILARITY: Contains 16 cell wall binding repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; D90213; BA14241.1; -  
DR InterPro: IPR002479; CW binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 13.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 38  
FT CHAIN 39 1592  
FT DOMAIN 39 1044  
FT DOMAIN 1093 1592  
FT REPEAT 1093 1592  
FT REPEAT 1093 1142  
FT REPEAT 1093 1207  
FT REPEAT 1222 1272  
FT REPEAT 1287 1337  
FT REPEAT 1402 1451  
FT REPEAT 1514 1563  
FT REPEAT 1577 1592  
SQ SEQUENCE 1592 AA; 176167 MW; BCOA66D079351ECF CRC64;  
Query Match 62.7%; Score 69; DB 1; Length 1592;  
Best Local Similarity 68.2%; Pred. No. 0.0044;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

RESULT 7  
SYD\_TREPA STANDARD; PRT; 602 AA.  
ID SYD\_TREPA  
AC O83950;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate--tRNA ligase)  
OS (Aspergillus).  
GN ASPS OR TP0985.  
OC Treponema pallidum.  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
CX NCBI\_TaxID=160;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nichols;  
RX MEDLINE=9832770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ueberback T.,  
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete.";  
RL Science 281:375-388(1998).  
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +  
CC diphosphate + L-aspartyl-tRNA(Asp).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AB001266; AAC65942.1; -  
DR PIR; A71256; A71256.  
DR HSSP; P36419; IEFW.  
DR TIGR; TP0985; -  
DR HAMAP; MF 00044; -; 1.  
DR InterPro: IPR004524; AspS\_bact.  
DR InterPro: IPR004115; GAD\_dom.  
DR InterPro: IPR008994; Nucleic\_acid\_OB.  
DR InterPro: IPR004364; tRNA-synt\_2.  
DR InterPro: IPR002312; tRNA-synt\_2.  
DR InterPro: IPR004365; tRNA-anti\_1.  
DR InterPro: IPR006195; tRNA\_ligase\_II.  
DR Pfam; PF02938; GAD; 1.  
DR Pfam; PF00152; tRNA-synt\_2; 2.  
DR Pfam; PF01336; tRNA-anti\_1.  
DR PRINTS; PR01042; TRNASYNTHASP.  
DR TIGRFAMs; TIGR00459; aspS\_bact; 1.  
DR PROSITE; PSS0862; AA\_tRNA\_LIGASE\_II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 602 AA; 68626 MW; AS1D0B17DE4FC5BC CRC64;  
Query Match 44.5%; Score 49; DB 1; Length 602;  
Best Local Similarity 56.2%; Pred. No. 3.1;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

RESULT 8
RA51 DROME STANDARD; PRT; 336 AA.
AC 027257; G9VAA8.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA repair protein Rad51 homolog (RecA protein homolog).
GN RAD51 OR DMR OR CG7948.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Canton-S;
RC MEDLINE=95161094; PubMed=7857671;
RA Akaboshi E., Inoue Y., Ryo H.;
RT "Cloning of the cDNA and genomic DNA that correspond to the recA-like
RT gene of Drosophila melanogaster.";
RL Upn. J. Genet. 69:663-670 (1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Oregon-R;
RC MEDLINE=96207535; PubMed=8625736;
RA McKee B.D., Ren X.J., Hong C.S.;
RT "A recA-like gene in Drosophila melanogaster that is expressed at
RT high levels in female but not male meiotic tissues.";
RL Chromosoma 104:479-488 (1996).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Aeshbuner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champagne B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley B.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan F., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostali M., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Peltzman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
-1- FUNCTION: Binds to single and double stranded DNA and exhibits

```

```

CC DNA-dependent ATPase activity. Underwinds duplex DNA (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Belongs to the recA family. Rad51 subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D37788; BAA07039.1; -
DR EMBL; D17726; BAA04580.1; -
DR EMBL; L41342; AAA64873.1; -
DR EMBL; AE003772; AAF57005.1; -
DR HSSR; 006609; 1B22.
DR P1Ybase; FBgn011700; Rad51.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000445; HnH.
DR InterPro; IPR003583; HnH.
DR Pfam; PF00633; HnH; 1.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00278; HnH; 1.
DR PROSITE; PS50162; RECA_2; 1.
DR PROSITE; PS50163; RECA_3; 1.
KW DNA-binding; ATP-binding; Nuclear protein.
NP BIND 124 131
SQ SEQUENCE 336 AA; 36647 MW; F9B9B21405B15DB0 CRC64;

```

```

Query Match 42.7%; Score 47; DB 1; Length 336;
Best Local Similarity 42.1%; Pred. No. 3.6;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Cy 3 MYVIRANDESVQRIAKI 21
Db 185 NVAFTRAHNSDQTKLIOM 203

RESULT 9
SYD MYCPN STANDARD; PRT; 557 AA.
AC P75068;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate--tRNA ligase)
DE (AspRS).
GN ASPS OR MPN046 OR MP108.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2104;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 29342 / M229;
RC MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +
CC diphosphate + L-aspartyl-tRNA(Asp).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE000013; AAB95756.1; -  
DR PIR: S73434; S73434.  
DR HSSP: P36419; 1EFW.  
DR HAMAP: MF\_00044; -; 1.  
DR InterPro: IPR004524; Asps bact.  
DR InterPro: IPR008994; Nucleic acid\_OB.  
DR InterPro: IPR004364; tRNA-synt 2.  
DR InterPro: IPR002312; tRNA-synt asp.  
DR InterPro: IPR004365; tRNA anti.  
DR InterPro: IPR006195; tRNA\_ligase\_II.  
DR Pfam: PF00152; tRNA-synt\_2; 2.  
DR Pfam: PF01336; tRNA anti; 1.  
DR PRINTS: PRO1042; TRNASYNTHASP.  
DR TIGRPFAM: TIGR00459; asps bact; 1.  
DR PROSITE: PS50862; AA\_TRNA\_LIGASE\_II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 557 AA; 64114 MW; 65EBE73A85F8A424 CRC64;

Query Match 42.7%; Score 47; DB 1; Length 557;  
Best Local Similarity 75.0%; Pred. No. 6.1;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 IRAHSEVQTRI 18  
DB 463 IRIHDEVTQL 474  
|||:|||||  
|||:|||||

RESULT 10  
SYD BORB  
ID SYD BORB STANDARD; PRT; 586 AA.  
AC OS1402;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate-tRNA ligase)  
DE (Asprs).  
GN ASPS OR BB0446.  
OS *Borrelia burgdorferi* (Lyme disease spirochete).  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; *Borrelia*.  
OX NCBI\_TaxID=139;  
RX STRAIN=ATCC 35210 / B31;  
RP MEDLINE=98065943; Pubmed=9403685.  
RA Fraser C.M., Caeyens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kervase A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utechack T., White L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "genomic sequence of a Lyme disease spirochete, *Borrelia*  
RT *burgdorferi*.";  
RT Nature 390:580-586(1997).  
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +  
CC diphosphate + L-aspartyl-tRNA(Asp).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE01149; AAB91506.1; ALT\_INIT.  
DR PIR: E70155; E70155.  
DR HSSP: P36419; 1EFW.  
DR TIGR: BB0446; -  
DR HAMAP: MF\_00044; -; 1.  
DR InterPro: IPR004524; Asps bact.  
DR InterPro: IPR004115; GAD dom.  
DR InterPro: IPR008994; Nucleic acid\_OB.  
DR InterPro: IPR004364; tRNA-synt 2.  
DR InterPro: IPR002312; tRNA-synt asp.  
DR InterPro: IPR004365; tRNA anti.  
DR InterPro: IPR006195; tRNA\_ligase\_II.  
DR Pfam: PF02938; GAD; 1.  
DR Pfam: PF00152; tRNA-synt\_2; 2.  
DR Pfam: PF01336; tRNA anti; 1.  
DR PRINTS: PRO1042; TRNASYNTHASP.  
DR TIGRPFAM: TIGR00459; asps bact; 1.  
DR PROSITE: PS50862; AA\_TRNA\_LIGASE\_II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 586 AA; 68343 MW; B7BADE0C44BE16EB CRC64;

Query Match 42.7%; Score 47; DB 1; Length 586;  
Best Local Similarity 62.5%; Pred. No. 6.4;  
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 IRAHSEVQTRI 22  
DB 486 IRIHKEIQRIKII 501  
|||:|||||  
|||:|||||

RESULT 11  
SYD THERN  
ID SYD THERN STANDARD; PRT; 592 AA.  
AC OSR17;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate-tRNA ligase)  
DE (Asprs).  
GN ASPS OR TTE1231.  
OS *Thermotoga tengcongensis*.  
OC Bacteria; Firmicutes; Clostridia; Thermotoga bacteriales;  
OC Thermotogae; Thermotogaceae; Thermotogaceae.  
OX NCBI\_TaxID=119072;  
RX STRAIN=MB4 / JCM 11007;  
RP MEDLINE=21992816; Pubmed=11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of *T. tengcongensis* genome.";  
RT Genome Res. 12:689-700(2002).  
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +  
CC diphosphate + L-aspartyl-tRNA(Asp).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE013085; AAM24458.1; -  
DR HAMAP: MF\_00044; -; 1.  
DR InterPro: IPR004524; Asps bact.  
DR InterPro: IPR004115; GAD dom.

DR InterPro: IPR008994; Nucleic acid OB.  
 DR InterPro: IPR004364; tRNA-synt 2.  
 DR InterPro: IPR002312; tRNA-synt asp.  
 DR InterPro: IPR004365; tRNA anti.  
 DR InterPro: IPR006195; tRNA\_ligase\_II.  
 DR Pfam: PF02938; GAD; 1.  
 DR Pfam: PF00152; tRNA-synt 2; 2.  
 DR Pfam: PF01336; tRNA anti; 1.  
 DR PRINTS: PR01042; tRNA SYNTHASP.  
 DR TIGRFAMs: TIGR00459; aaps\_bact; 1.  
 DR PROSITE: PS50862; AA tRNA\_LIGASE\_II; 1.  
 DR AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KM Complete proteome.  
 SQ SEQUENCE 592 AA; 67864 MW; 1621BA9C6BC30381 CRC64;

Query Match 42.7%; Score 47; DB 1; Length 592;  
 Best Local Similarity 50.0%; Pred. No. 6.5;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 7 IRAHDSFVOTRIAKI 22  
 DB 495 IRIHDTLQCKMKVYL 510

RESULT 12  
 CBP1\_YEAST STANDARD; PRT; 654 AA.

ID CBP1\_YEAST STANDARD; PRT; 654 AA.  
 AC P07252;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Cytochrome B pre-mRNA processing protein 1.  
 GN CBP1 OR YUL209W OR J0242 OR HRA654.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;

[1]  
 SEQUENCE FROM N.A.  
 MEDLINE=84185566; PubMed=6325407;  
 RA Dieckmann C.L., Homlson G., Tagoloff A.;  
 RT "Assembly of the mitochondrial membrane system. Nucleotide sequence  
 of a yeast nuclear gene (CBP1) involved in 5' end processing of  
 cytochrome b pre-mRNA.";  
 RL J. Biol. Chem. 259:4732-4738 (1984).

[2]  
 SEQUENCE FROM N.A.  
 STRAIN=S288C;  
 RC MEDLINE=95242842; PubMed=7725802;  
 RA Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,  
 RA Hilger F.;  
 RT "Sequence analysis of a 40.2 kb DNA fragment located near the left  
 telomere of yeast chromosome X.";  
 RL Yeast 10:1657-1662 (1994).

[3]  
 SEQUENCE OF 589-654 FROM N.A.  
 MEDLINE=90014786; PubMed=2552292;  
 RA Liu Y., Dieckmann C.L.;  
 RT "Overproduction of yeast viruslike particles by strains deficient in  
 a mitochondrial nuclease.";  
 RL Mol. Cell. Biol. 9:3323-3331 (1989).  
 -1- FUNCTION: Responsible for conferring a stable 5' end on cytochrome  
 b mRNA.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; KO2647; AAA34474.1; -

DR EMBL: Z34098; CAA84002.1; -  
 DR EMBL: Z49484; CAA89506.1; -  
 DR EMBL: M28067; AAA34456.1; -  
 DR PIR: S05829; BVBVPL.  
 DR GerMOnline: 141821; -  
 DR SGD: S0003745; CBP1.  
 DR GO: GO:0005739; Mitochondrion; IDA.  
 DR GO: GO:0003729; rRNA binding; IGI.  
 DR GO: GO:0009060; P aerobic respiration; IMP.  
 DR GO: GO:0006402; P mRNA catabolism; IMP.  
 KM mRNA processing.  
 SQ SEQUENCE 654 AA; 76171 MW; 2453B03280E1C44D CRC64;

Query Match 40.9%; Score 45; DB 1; Length 654;  
 Best Local Similarity 47.4%; Pred. No. 15;  
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSFVOTRIAKI 21  
 DB 300 HYVSIHDSRIRITYKVKFI 318

RESULT 13

AROA\_PYRFU STANDARD; PRT; 406 AA.

ID AROA\_PYRFU STANDARD; PRT; 406 AA.  
 AC Q8U0A0;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)  
 DE (5-enolpyruvylshikimate-3-phosphate synthase) (EPSPS).  
 GN AROA OR PF1659.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2261;

[1]  
 SEQUENCE FROM N.A.  
 STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RT "The complete sequence of the Pyrococcus furiosus genome.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =  
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.  
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
 CC sixth step.

CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: AE010267; AAL81823.1; ALT\_INIT.  
 DR HAMAP: MF\_00210; -; 1.  
 DR InterPro: IPR006264; AROA.  
 DR InterPro: IPR001986; EPSP synth.  
 DR Pfam: PF00275; EPSP\_synthase; 1.  
 DR ProDom: PD001867; EPSP\_synthase; 1.  
 DR TIGRFAMs: TIGR01356; aroA; 1.  
 DR PROSITE: PS00104; EPSP SYNTHASE 1; FALSE\_NEG.  
 DR PROSITE: PS00885; EPSP SYNTHASE 2; 1.  
 KM Aromatic amino acid biosynthesis; Transferrase; Complete proteome.  
 SQ SEQUENCE 406 AA; 45068 MW; 43AACD7FE97C272 CRC64;

Query Match 39.1%; Score 43; DB 1; Length 406;  
 Best Local Similarity 47.6%; Pred. No. 20;  
 Matches 10; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 2 PNVVFIKADSEVQTRIAKII 22  
 DB 75 PNVIF--ARDSGTARISITIV 93

## RESULT 14

ARGD BIFLO STANDARD; PRT; 431 AA.  
 AC PS9315;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Acetylornithine aminotransferase (EC 2.6.1.11) (ACOA)  
 GN ARGD OR BL1061  
 OS Bifidobacterium longum.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;  
 OC Bifidobacteriaceae; Bifidobacterium.  
 OX NCBI\_TaxID=216816;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NC2705;  
 RX MEDLINE=22294977; PubMed=12381787;  
 RA Schell M.A., Karmaliantzou M., Snel B., Vilanova D., Berger B.,  
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,  
 RA Pridmore R.D., Arigoni F.;  
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation  
 to the human gastrointestinal tract."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
 CC -1- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + 2-oxoglutarate = N-  
 acetyl-L-glutamate 5-semialdehyde + L-glutamate.  
 CC -1- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -1- PATHWAY: Arginine biosynthesis; fourth step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- MISCELLANEOUS: May also have succinyl-L-tyrosinyl-L-lysine  
 aminotransferase activity, thus carrying out the fourth step in  
 lysine biosynthesis.  
 CC -1- SIMILARITY: Belongs to class-III of pyridoxal-phosphate-dependent  
 aminotransferases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE014729; AAN24869.1; -.  
 DR HAMAP; MF 01107; -; 1.  
 DR InterPro; IPR005814; AminoTrans\_3.  
 DR InterPro; IPR004636; ARGD.  
 DR Pfam; PF00202; aminoTrans\_3; 1.  
 DR TIGRFAMs; TIGR00707; argD; 1.  
 DR PROSITE; PS00600; AA\_TRANSFER\_CLASS\_3; 1.  
 KM Arginine biosynthesis; Transferase; Aminotransferase;  
 KM Pyridoxal phosphate; Complete proteome.  
 FT BINDING 284  
 FT BINDING 284 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 431 AA; 45615 MW; 1A9245B10351BCE CRC64;

Query Match 39.1%; Score 43; DB 1; Length 431;  
 Best Local Similarity 38.5%; Pred. No. 21;  
 Matches 10; Conservative 4; Mismatches 6; Indels 6; Gaps 1;

QY 1 PNVVFIKADSEVQTRIAK 20  
 DB 181 PNVIFVRAADKALHDAFAQTGIGR 206

RESULT 15  
 YMDL IACIA STANDARD; PRT; 319 AA.  
 AC Q9CG88; Q34131;

DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein UPF0176 protein ymdL  
 GN YMDL OR LI1217.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobactillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NC20118;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Winkler P., Manger S., Jallion O., Malarme K.,  
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 lactis sep. lactis IL1403."  
 RL Genome Res. 11:731-753(2001).  
 CC -1- SIMILARITY: Belongs to the UPF0176 family.  
 CC -1- SIMILARITY: Contains 1 rhodanese domain.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U92974; AAB81912.1; -.  
 DR EMBL; AB06353; AAK05315.1; -.  
 DR PIR; A86777; A86777.  
 DR HAMAP; MF 00469; -; 1.  
 DR InterPro; IPR001763; Rhodanese-like.  
 DR Pfam; PF00581; Rhodanese; 1.  
 DR SMART; SM00450; RHOD; 1.  
 DR PROSITE; PS50206; RHODANES\_3; 1.  
 KM Hypothetical protein; Complete proteome.  
 FT DOMAIN 125  
 FT DOMAIN 125 RHODANES.  
 FT CONFICT 204  
 FT CONFICT 204 H -> L (TN REF. 1).  
 SQ SEQUENCE 319 AA; 36849 MW; F987D509415F169A CRC64;

Query Match 38.6%; Score 42.5; DB 1; Length 319;  
 Best Local Similarity 47.6%; Pred. No. 19;  
 Matches 10; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

QY 2 PNVVFIKADSEVQTRIA 19  
 DB 291 PNNRYIKAHQLSNOEVOERLA 311

RESULT 16  
 Y943 MYCTU STANDARD; PRT; 346 AA.  
 AC P71566;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein RV0943C/MT0965/MD0966C.  
 GN RV0943C OR MT0969 OR MTCY10D7.31 OR MB0966C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773; 1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;



```

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Biglieri K., Gas S., Barry C.E. III, Tekla F.,
RA Badcock K., Baaham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagsen K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22306494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gilm M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=2279107; PubMed=12788972;
RA Garnier T., Biglieri K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsepe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z79700; CAB01978.1; -.
DR EMBL: AB006982; AAK45217.1; ALT_INT.
DR EMBL: BX248337; CAD93829.1; -.
DR PIR: E70715; E70715.
DR TIGR: MT0969; -.
DR Tuberculist: RV0943c; -.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 37205 MW; F51A58155A36136 CRC64;

```

Query Match 38.2%; Score 42; DB 1; Length 346;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

QY 2 PNYVPIRAHSEVQTR 17
Db 177 PNYFFITGPTTAQAR 192

```

RESULT 17  
DPO4\_SULTO STANDARD; PRT; 353 AA.  
AC Q974T8;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA polymerase IV (EC 2.7.7.7) (Pol IV).  
GN DBH OR ST0573.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.

```

OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anhal A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takemura M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagita M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7."
RL DNA Res. 8:123-140 (2001).
CC -----
CC -1- FUNCTION: Poorly processive, error-prone DNA polymerase involved
CC in untargeted mutagenesis. Copies undamaged DNA at stalled
CC replication forks, which arise in vivo from mismatched or
CC misaligned primer ends. These misaligned primers can be extended
CC by pol IV. Exhibits no 3'-5' exonuclease (proofreading) activity.
CC May be involved in translational synthesis (by similarity).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- COFACTOR: Binds 2 magnesium ions per subunit (by similarity).
CC -1- SIMILARITY: Belongs to the DNA polymerase type-X family.
CC -1- SIMILARITY: Contains 1 umuc domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000983; BAB65569.1; -.
DR HAMAP: MF_01113; -.
DR Interpro: IPR001126; UMOUC_1like.
DR Pfam: PF00817; IMS; 1.
DR PROSITE: PS50173; UMOUC; 1.
KM Transference; DNA-directed DNA polymerase; DNA replication; DNA repair;
KM DNA-binding; Mutator protein; Magnesium; Complete proteome.
FT DOMAIN 5 183 UMOUC
FT SITE 14 14 SUBSTRATE DISCRIMINATION (BY SIMILARITY).
FT ACT SITE 108 108 BY SIMILARITY.
FT METAL 9 9 MAGNESIUM (BY SIMILARITY).
FT METAL 107 107 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 353 AA; 40125 MW; 1C3B46ACDC98D55 CRC64;

```

Query Match 37.7%; Score 41.5; DB 1; Length 353;  
Best Local Similarity 41.7%; Pred. No. 30;  
Matches 10; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

```

QY 2 PNYVPIRAHSEVQTR 22
Db 71 PNAIFVMEKXVYTVSNRIMSI 94

```

RESULT 18  
GREX\_MYCPN STANDARD; PRT; 160 AA.  
AC P78019;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Transcription elongation factor grex (Transcript cleavage factor  
DE grex).  
GN GREX OR MPN401 OR MP437.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;



```

RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreuch R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- FUNCTION: Necessary for efficient RNA polymerase transcription
CC elongation past template-encoded arresting sites. The arresting
CC sites in DNA have the property of trapping a certain fraction of
CC elongating RNA polymerases that pass through, resulting in locked
CC ternary complexes. Cleavage of the nascent transcript by cleavage
CC factors such as greA or greB allows the resumption of elongation
CC from the new 3' terminus. GreA releases sequences of 2 to 3
CC nucleotides (By similarity).
CC -1- SIMILARITY: Belongs to the greA/greB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000043; AAB96085.1; -.
DR PIR; S73763; S73763.
DR HSSP; P21346; IGRJ.
DR HAMAP; MF_00105; -.
DR InterPro; IPR006359; GreA.
DR InterPro; IPR001437; GreA_Greb.
DR Pfam; PF01272; GreA_Greb; 1.
DR ProDom; PD004918; GreA_Greb_N; 1.
DR TIGRFAMs; TIGR01462; greA; 1.
DR PROSITE; PS00829; GREAB_2; 1.
DR PROSITE; PS00830; GREAB_2; 1.
DR Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 8 28 COILED COIL (POTENTIAL).
FT DOMAIN 48 75 COILED COIL (POTENTIAL).
SQ SEQUENCE 160 AA; 18101 MW; 32448C91712A19C2 CRC64;

Query Match 37.3%; Score 41; DB 1; Length 160;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIKADSEVOTRIAKI 21
DB 49 DYDAKAKOGGEIRIRIAEI 67

RESULT 19
GREA_MYCSE STANDARD; PRT; 161 AA.
ID GREA_MYCSE
AC P47524;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor
DE greA).
GN GREA OR MG282.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_Taxid=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=5569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

```

```

RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -1- FUNCTION: Necessary for efficient RNA polymerase transcription
CC elongation past template-encoded arresting sites. The arresting
CC sites in DNA have the property of trapping a certain fraction of
CC elongating RNA polymerases that pass through, resulting in locked
CC ternary complexes. Cleavage of the nascent transcript by cleavage
CC factors such as greA or greB allows the resumption of elongation
CC from the new 3' terminus. GreA releases sequences of 2 to 3
CC nucleotides (By similarity).
CC -1- SIMILARITY: Belongs to the greA/greB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39708; AAC71504.1; -.
DR PIR; B64231; B64231.
DR HSSP; P21346; IGRJ.
DR TIGR; MG282; -.
DR HAMAP; MF_00105; -.
DR InterPro; IPR006359; GreA.
DR InterPro; IPR001437; GreA_Greb.
DR Pfam; PF01272; GreA_Greb; 1.
DR ProDom; PD004918; GreA_Greb_N; 1.
DR TIGRFAMs; TIGR01462; greA; 1.
DR PROSITE; PS00829; GREAB_2; 1.
DR PROSITE; PS00830; GREAB_2; 1.
DR Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 9 28 COILED COIL (POTENTIAL).
FT DOMAIN 47 76 COILED COIL (POTENTIAL).
SQ SEQUENCE 161 AA; 18162 MW; 67E2850CB59BCC5B CRC64;

Query Match 37.3%; Score 41; DB 1; Length 161;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIKADSEVOTRIAKI 21
DB 49 DYDAKAKOGGEIRIRIAEI 67

RESULT 20
KADA_MOUSE STANDARD; PRT; 223 AA.
ID KADA_MOUSE
AC O9WUE9; O9RLX7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adenylate kinase isoenzyme 4, mitochondrial (EC 2.7.4.3) (ATP-AMP
DE transphosphorylase).
GN AK4 OR AK-4 OR AK3B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99033072; PubMed=9813319;
RA Yoneda T., Sato M., Maeda M., Takagi H.;
RT "Identification of a novel adenylate kinase system in the brain:
RT cloning of the fourth adenylate kinase."
RL Brain Res. Mol. Brain Res. 62:187-195(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Noma T.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

```



```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SSI-1 / Serotype M3:
RX MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurikawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-MGAS822 / Serotype M18:
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbhan K.D., Van Gompel J.U., Smoot L.M., Chaussee M.S.,
RA Sylla G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Yeary L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreak.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC -1 SIMILARITY: Belongs to the Adomet synthetase 2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006511; AAK3533.1; -
CC EMBL: AE014144; AAM78989.1; -
CC EMBL: AP005145; BAC64566.1; -
CC EMBL: AE009966; AAL97295.1; -
CC HAMAP: MF_00136; atypical; 1.
CC InterPro: IPR002795; MAT.
CC Pfam: PF01941; Adomet_Synthetase; 1.
CC DR Prodom: PD011777; MAT; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 399 AA; 45771 MW; 590FC44690960103 CRC64;

```

```

Query Match 37.3%; Score 41; DB 1; Length 399;
Best Local Similarity 33.3%; Pred. No. 42;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 2 PNYVIRADSEVOTRIAKII 22
DB 350 PTHLFOTSESVDERVRLIV 370

```

```

RESULT 23
SYD_BUCAP STANDARD; PRT; 584 AA.
AC P81432;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate--tRNA ligase)
DE (Asp88).
GN ASPS OR BUSG306.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RA Thao M.L., Baumann P.;
RT "Nucleotide sequence of a DNA fragment from Buchnera aphidicola
RT (Aphid endosymbiont) containing the genes asps-tRNA-ber-se-c-aro-a-
RT tpaA-hmd-tpaA.";

```

```

RL Curr. Microbiol. 35:68-69(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tanas I., Klassen L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Mennergreen J.J., Sanderrom J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -1 CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +
CC dihydrophosphate + L-aspartyl-tRNA(Asp).
CC -1 SUBUNIT: Homodimer (By similarity).
CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1 SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L43549; AAC05432.1; -
CC EMBL: AB014107; AAM67860.1; -
CC HSRP; P21889; 1BOR.
CC HAMAP; MF_00044; -; 1.
CC InterPro: IPR004524; ASPS_bact.
CC InterPro: IPR004115; GAD_dom.
CC InterPro: IPR008994; Nucleic acid OB.
CC InterPro: IPR004364; tRNA-synt 2.
CC InterPro: IPR002312; tRNA-synt Asp.
CC InterPro: IPR004365; tRNA anti-
CC InterPro: IPR006195; tRNA_ligase_II.
CC Pfam; PF02938; GAD; 1.
CC Pfam; PF00152; tRNA-synt 2; 2.
CC Pfam; PF01336; tRNA anti; 1.
CC PRINTS; PR01042; TRNSYNTHASP.
CC TIGRPFAM; TIGR00459; asps_bact; 1.
CC PROSITE; PS50862; AA-tRNA_LIGASE_II; 1.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC KW Complete proteome.
CC FM DOMAIN 79
CC SQ SEQUENCE 584 AA; 67824 MW; F318DFA1654780E7 CRC64;

```

```

Query Match 37.3%; Score 41; DB 1; Length 584;
Best Local Similarity 37.5%; Pred. No. 62;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 7 IRADSEVOTRIAKII 22
DB 486 VRTHDKIKQKVFNII 501

```

```

RESULT 24
COG6_DROME STANDARD; PRT; 630 AA.
AC Q9V564; Q95RW7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative conserved oligomeric Golgi complex component 6.
GN CG1968.
OS Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```



```

FT DOMAIN 194 198 POLY-GIN.
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 565 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 718 718 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 737 AA; 83708 MW; B2850C1142F86C76 CRC64;

Query Match 37.3%; Score 41; DB 1; Length 737;
Best Local Similarity 42.9%; Pred. No. 78;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VENVYFIRADSEVQ 14
DB 555 MPNYDFLEHNSST 568

RESULT 26
VIA_BMV STANDARD; PRT; 961 AA.
AC P03588;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 1A protein [includes: Helicase; Methyltransferase].
OS Brome mosaic virus (BMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Bromovirus.
OC NCBI_TaxID=12302;
RX [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=84114904; PubMed=6694215;
RA Ahlquist P., Dasgupta R., Kaesberg P.;
RT "Nucleotide sequence of the brome mosaic virus genome and its
RT implications for viral replication."
RT J. Mol. Biol. 172:369-383(1984).
CC -1- FUNCTION: MAY BE INVOLVED IN THE REPLICATION OF THE VIRUS.
CC CONTAINS AN HELICASE DOMAIN AND A METHYLTRANSFERASE DOMAIN. THE
CC METHYLTRANSFERASE DOMAIN IS PROBABLY INVOLVED IN VIRAL RNA
CC CAPING.
CC -1- SIMILARITY: TO 1A PROTEIN FROM CCMV, CMV, PSV AND TAV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02380; CAA26228.1; -
CC PIR; A04196; PIBVA.
CC InterPro; IPR002588; V_methyltransf.
CC InterPro; IPR000606; Viral_helicase1.
CC Pfam; PF01443; Viral_helicase1.1.
CC Pfam; PF01660; Vmethyltransf.1.
CC KMW Helicase; ATP-binding; Transferase; Methyltransferase.
CC FT NP BIND 685 692 ATP (POTENTIAL).
CC FT SEQUENCE 961 AA; 109209 MW; 4F315CB2E2FAFBC CRC64;

Query Match 37.3%; Score 41; DB 1; Length 961;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 157 VENVYFIRADSEVQ 15
DB 157 VENVYFIRADSEVQ 171

RESULT 27
YA84 SCHPO STANDARD; PRT; 1428 AA.

```

```

AC 009773;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C22F3.04 in chromosome I.
GN SPAC22F3.04.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RX [1]
RX SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RX Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RX Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RX Collins M., Connor R., Cronin A., Davis P., Delwail T., Fraser A.,
RX Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RX Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,
RX James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RX Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RX Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RX Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RX Skelton J., Simmonds M., Squares R., Stevens K., Whitehead S.,
RX Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RX Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RX Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RX Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RX Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RX Eger P., Zimmermann W., Wedler H., Wambolt R., Punelie B.,
RX Goffeau A., Cadieu E., Dreano S., Gloux S., Leleu V., Motier S.,
RX Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RX Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RX Daga R.R., Cruzado L., Jimenez J., Moreno S., Armstrong J., Benteo J.,
RX Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Benteo J.,
RX Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RX Shpakovski G.V., Useery D., Barrell B.G., Nurre P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC -1- SIMILARITY: TO S.POMBE SPAC56F8.02.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z54285; CAA91069.1; -
CC PIR; T38187; S62419.
CC GenDB; Spombe; SPAC22F3.04; -
CC InterPro; IPR000873; AMP-bind.
CC Pfam; PF00501; AMP-binding; 3.
CC Hypothetical protein; Transmembrane.
CC KMW TRANSMEM 1028 1048 POTENTIAL.
CC FT TRANSMEM 1303 1323 POTENTIAL.
CC FT TRANSMEM 1333 1353 POTENTIAL.
CC FT SEQUENCE 1428 AA; 162394 MW; A88D03A80A9B9B9 CRC64;

Query Match 37.3%; Score 41; DB 1; Length 1428;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 8 RAHDSVQTRIAKI 21
DB 38 RAHDAELPOLAKV 51

RESULT 28

```

```

V224_FOMPV
ID V224_FOMPV STANDARD; PRT; 146 AA.
CC
CC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Putative ankryin-repeat protein FPV224.
GN FPV224.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC AVIPoxvirus.
OC NCBI_TaxID=10261;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Alonso C.L., Tulman E.R., Lu Z., Zaak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus."
RL J. Virol. 74:3815-3831(2000).
CC
CC -1- SIMILARITY: Contains 4 ANK repeats.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AF198100; AAF4568.1; -.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank. 3.
DR SMART: SMO0248; ANK_3
DR PROSITE; PSS0086; ANK_REPEAT; 1.
DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
KM Hypothetical protein; Repeat; ANK repeat.
FT REPEAT 9 38 ANK 1.
FT REPEAT 42 79 ANK 2.
FT REPEAT 94 126 ANK 3.
FT REPEAT 127 145 ANK 4.
SQ SEQUENCE 146 AA; 16759 MW; 9B1064458E4290B2 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 146;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYVFIRAHDS 12
Db 93 NYVFVNLHDA 102

RESULT 29
ADEN_ADEG8 STANDARD; PRT; 205 AA.
ID ADEN_ADEG8
AC O9QM72;
DT 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenahin (BC 3.4.22.39) (Endoprotease) (late L3 23 kDa protein).
OS Avian adenovirus type 8 (strain ATCC A-29) (Fowl adenovirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OC NCBI_TaxID=66295;
RN
RP SEQUENCE FROM N.A.
RA O'Kic D., Nagy E.;
RT "The DNA sequence of fowl adenovirus 8."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THIOLESTERASE SPECIFIC GLY-ALA PEPTIDES IN A
CC NUMBER OF VIRAL PRECURSOR PROTEINS (TERMINAL, ITA, VI, VII, VIII,
CC ITA AND MU) WHICH ARE REQUIRED FOR VIRION MATURATION. ALSO CLEAVES
CC HOST CELL CYTOSKELETAL KERATINS K7 AND K18.
CC -1- CATALYTIC ACTIVITY: Cleaves proteins of the adenovirus and its
CC host cell at two consensus sites: -Yaa-Xaa-Gly-Gly-I-Xaa- and
CC -Yaa-Xaa-Gly-Xaa-I-Gly- (in which Yaa is Met, Ile or Leu, and Xaa

```

```

CC is any amino acid).
CC -1- SIMILARITY: Belongs to peptidase family C5.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AF083975; AAD50345.2; -.
DR HSSP; P03252; IAVP.
DR MEROPS; C05.001; -.
DR InterPro: IPR000855; Peptidase_C5.
DR Pfam; PF00770; Peptidase_C5; 1.
DR PRINTS; PR00703; ADVENDOPTASE.
DR ProDom; PD003705; Peptidase_C5; 1.
KM Hydroxylase, thiol protease; Late protein.
FT ACT SITE 55 55 BY SIMILARITY.
FT ACT SITE 72 72 BY SIMILARITY.
FT ACT SITE 122 122 BY SIMILARITY.
SQ SEQUENCE 205 AA; 23701 MW; 36F0700CDF885F62 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 205;
Best Local Similarity 47.1%; Pred. No. 30;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQTRIA 19
Db 181 NNSFFRAHSESLKRETA 197

RESULT 30
SYD_MYCPE STANDARD; PRT; 580 AA.
ID SYD_MYCPE
AC O8EM87;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aspartyl-tRNA synthetase (BC 6.1.1.12) (Aspartate--tRNA ligase)
DE (AspRS).
GN ASPS OR MYP82870.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=28227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=1246555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans."
RL Nucleic Acids Res. 30:5293-5300(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA (Asp) = AMP +
CC diphosphate + L-aspartyl-tRNA (Asp).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AP004171; BAC44079.1; -.
DR HAMAP; MF_00044; -.
DR InterPro: IPR004524; APPS_bact.
DR InterPro: IPR004115; GAD_dom.

```

DR InterPro: IPR008994; Nucleic\_acid\_OR.  
DR InterPro: IPR004364; tRNA-synt\_2.  
DR InterPro: IPR002312; tRNA-synt\_ap.  
DR InterPro: IPR004365; tRNA anti.  
DR InterPro: IPR006195; tRNA\_ligase\_II.  
DR Pfam: PF02938; GAD; 1.  
DR Pfam: PF00152; tRNA-synt\_2; 2.  
DR Pfam: PF01336; tRNA anti; 1.  
DR PRINTS: PR01042; TRNtASYNTHASP.  
DR TIGRFAMs: TIGR00459; aaps\_bact; 1.  
DR PROSITE: PS0862; AA\_tRNA\_ligase\_II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding  
KW Complete proteome.  
SQ SEQUENCE 580 AA; 67119 MW; D13AR0DCD4741313 CRC64;

Query Match	36.4%	Score 40	DB 1	Length 580
Best Local Similarity	43.8%	Pred. No. 89		
Matches	7	Conservative	6	Mismatches 3
				Indels 0
				Gaps 0

```
QY      7 IRAHDSEVQTRIAKII 22  
        :|::||:||||:|:  
Db      483 LRIYNSELQTRMFKEI 498
```

Search completed: May 4, 2004, 09:10:41  
Job time : 11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 4, 2004, 09:05:56 ; Search time 33 Seconds  
(Without alignments)  
210.345 Million cell updates/sec

Title: US-09-290-049a-18  
Perfect score: 110  
Sequence: 1 VPMYFIRAHSEVQTRAKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapect 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	110	100.0	Q9WJ4	Q9WJ4 streptococc
2	90	81.8	Q84CN4	Q84CN4 leuconostoc
3	90	81.8	Q91466	Q91466 leuconostoc
4	90	81.8	Q9EZH5	Q9EZH5 leuconostoc
5	90	81.8	Q52224	Q52224 leuconostoc
6	90	81.8	Q8G9G2	Q8G9G2 leuconostoc
7	89	80.9	Q9ZAR4	Q9ZAR4 leuconostoc
8	89	80.9	Q8KRE1	Q8KRE1 leuconostoc
9	88	80.0	Q9WJ5	Q9WJ5 streptococc
10	87	79.1	Q91CJ7	Q91CJ7 leuconostoc
11	87	79.1	Q8KZL5	Q8KZL5 streptococc
12	86	78.2	Q48756	Q48756 leuconostoc
13	83	75.5	Q91CH3	Q91CH3 streptococc
14	83	75.5	Q54178	Q54178 streptococc
15	77	70.0	Q55265	Q55265 streptococc
16	75	68.2	Q00599	Q00599 streptococc

17	74	67.3	1449	2	Q68542	Q68542 streptococc
18	73	67.3	1449	2	Q55264	Q55264 streptococc
19	73	66.4	1518	2	Q00600	Q00600 streptococc
20	69	62.7	1590	2	Q59983	Q59983 streptococc
21	69	62.7	1590	2	Q55263	Q55263 streptococc
22	63	57.3	2057	2	Q9PR05	Q9PR05 leuconostoc
23	50	45.5	1442	10	Q9XHV5	Q9XHV5 oryza sativ
24	48	43.6	492	16	Q9WJX1	Q9WJX1 thermotoga
25	47	42.7	279	5	Q8IMJ5	Q8IMJ5 drosophila
26	47	42.7	284	5	Q8IGG8	Q8IGG8 drosophila
27	45.5	41.4	783	5	Q8SX46	Q8SX46 drosophila
28	45.5	41.4	498	16	Q9A9H5	Q9A9H5 callobacter
29	45	40.9	639	2	Q9AQR3	Q9AQR3 bacillus sp
30	45	40.9	245	5	Q817Y7	Q817Y7 tetrahymus
31	44.5	40.5	771	16	Q8EDZ6	Q8EDZ6 shewanella
32	44.5	40.5	111	16	Q8XLB3	Q8XLB3 clostridium
33	44	40.0	455	17	Q974G8	Q974G8 sulfolobus
34	44	40.0	583	10	Q45505	Q45505 arabidopsis
35	44	40.0	592	16	Q7UFY6	Q7UFY6 rhodospirillum rubrum
36	44	40.0	631	16	Q9BPL8	Q9BPL8 mycoplasma
37	44	40.0	741	2	Q810X8	Q810X8 chlorobium
38	44	40.0	790	16	Q8P986	Q8P986 xanthomonas
39	44	40.0	323	2	Q83XH6	Q83XH6 vibrio angu
40	43.5	39.5	397	16	Q8AAT1	Q8AAT1 bacteroides
41	43.5	39.5	727	10	Q8EXG9	Q8EXG9 physcomitres
42	43.5	39.5	727	10	Q948Q6	Q948Q6 physcomitres
43	43.5	39.5	210	5	Q9NGL5	Q9NGL5 leishmania
44	43	39.1	280	16	Q81VM8	Q81VM8 bacillus an
45	43	39.1	339	16	Q8A756	Q8A756 bacteroides
46	43	39.1	407	5	Q88YP9	Q88YP9 drosophila
47	43	39.1	407	10	Q94GN4	Q94GN4 oryza sativ
48	43	39.1	457	16	Q9BPK4	Q9BPK4 chlamydia m
49	43	39.1	457	16	Q84126	Q84126 chlamydia t
50	43	39.1	547	12	Q8GY02	Q8GY02 norwalk-like
51	43	39.1	692	16	Q8RAS7	Q8RAS7 thermococcus
52	43	39.1	765	16	P74599	P74599 synchocyst
53	43	39.1	1070	16	Q55365	Q55365 synchocyst
54	43	39.1	2879	5	Q9U0Z1	Q9U0Z1 leishmania
55	43	39.1	164	16	Q8BOA6	Q8BOA6 mus musculus
56	42	38.2	148	11	Q9XAL5	Q9XAL5 bacillus ha
57	42	38.2	207	2	Q91650	Q91650 lactococcus
58	42	38.2	261	5	Q7YX07	Q7YX07 caenorhabditis
59	42	38.2	261	5	Q81U75	Q81U75 bacillus ce
60	42	38.2	321	17	Q8TNU3	Q8TNU3 methanobac
61	42	38.2	394	2	Q8RQW5	Q8RQW5 chitinophag
62	42	38.2	481	2	Q9AOL4	Q9AOL4 chitinophag
63	42	38.2	481	2	Q9PAX0	Q9PAX0 flexibacter
64	42	38.2	501	5	Q964R1	Q964R1 blattella g
65	42	38.2	501	5	Q964R1	Q964R1 drosophila
66	42	38.2	520	5	Q9W5W9	Q9W5W9 drosophila
67	42	38.2	587	16	Q8A9E3	Q8A9E3 bacteroides
68	42	38.2	915	3	Q96UB2	Q96UB2 neurospora
69	42	38.2	1534	13	Q8UG61	Q8UG61 brachydanio
70	42	38.2	2206	12	Q83017	Q83017 lactate deh
71	41.5	37.7	95	17	Q8UIG5	Q8UIG5 pyrococcus
72	41	37.3	122	16	Q983W0	Q983W0 rhizobium 1
73	41	37.3	122	16	Q94G24	Q94G24 physalis 10
74	41	37.3	218	2	Q93C95	Q93C95 campylobact
75	41	37.3	222	16	Q88AY6	Q88AY6 pseudomonas
76	41	37.3	233	2	Q9KSC9	Q9KSC9 campylobact
77	41	37.3	239	16	Q7UUX5	Q7UUX5 rhodospirillum rubrum
78	41	37.3	245	9	Q81TH1	Q81TH1 streptococcus
79	41	37.3	245	16	Q9S8S7	Q9S8S7 streptococcus
80	41	37.3	265	3	Q9C2G8	Q9C2G8 neurospora
81	41	37.3	270	2	Q54358	Q54358 streptomyces
82	41	37.3	282	10	Q9WJX2	Q9WJX2 arabidopsis
83	41	37.3	282	16	Q87ZB6	Q87ZB6 pseudomonas
84	41	37.3	292	16	Q82VZ6	Q82VZ6 nitrosomonas
85	41	37.3	294	16	Q8DUZ7	Q8DUZ7 streptococcus
86	41	37.3	325	16	Q7V142	Q7V142 prochlorococcus
87	41	37.3	353	13	Q93247	Q93247 cyprinus ca
88	41	37.3	389	3	Q872N5	Q872N5 cyprinus ca
89	41	37.3	407	12	Q65173	Q65173 african swi



90 41 37.3 416 2 Q9XBQ8  
 91 41 37.3 457 16 Q88RR4  
 92 41 37.3 451 16 Q88P89  
 93 41 37.3 484 16 Q89VC8  
 94 41 37.3 488 16 Q89VC8  
 95 41 37.3 560 4 Q14520  
 96 41 37.3 609 10 Q9JUC5  
 97 41 37.3 688 5 Q45430  
 98 41 37.3 712 10 Q93WP2  
 99 41 37.3 712 10 Q8L435  
 100 41 37.3 805 5 Q9TVL8

## ALIGNMENTS

RESULT 1  
 Q9WKT4 PRELIMINARY; PRT; 1338 AA.

AC Q9WKT4; 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE GTP-8.  
 GN GTP-8.  
 OS Streptococcus criceti.  
 OG Plasmid PAM.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS-6.  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S.citaceus glucosyltransferase(gtfs and gtf) genes";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB026123; BAA77236.1; -  
 DR GO: GO:0046821; C:extrachromosomal DNA; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CM binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CM binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM PfamId.  
 SQ SEQUENCE 1338 AA; 148558 MW; 0A90CE10B15D99B CRC64;

Query Match 100.0%; Score 110; DB 2; Length 1338;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENVYFIRAHDSVQVTRIAKII 22  
 DB 509 VENVYFIRAHDSVQVTRIAKII 530

RESULT 2  
 Q84CN4 PRELIMINARY; PRT; 1330 AA.

AC Q84CN4; 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Dextranucrase Dsrr (EC 2.4.1.5).  
 GN Dsrr.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1501;  
 RA Kim C.H., Moon J.O., Jang E.K.;  
 RT "Gene encoding a dextranucrase (Dsrr) in Leuconostoc mesenteroides  
 NRRL B-1501.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY142210; AAN38835.1; -  
 DR GO: GO:0016757; P:transferase activity, transferring glycosyl...; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CM binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CM binding\_1; 11.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1330 AA; 148863 MW; D945CB36CF75797 CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1330;  
 Best Local Similarity 77.3%; Pred. No. 1.9e-06;  
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VENVYFIRAHDSVQVTRIAKII 22  
 DB 456 IENVYFIRAHDSVQVTRIAKII 477

RESULT 3

Q9L466 PRELIMINARY; PRT; 1477 AA.  
 AC Q9L466; 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Dextranucrase (EC 2.4.1.5).  
 GN Dsrr.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1355;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,  
 RA Willems R.M., Monsan P.;  
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
 RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A250172; CAB76565.1; -  
 DR GO: GO:0016757; P:transferase activity, transferring glycosyl...; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CM binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CM binding\_1; 13.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1477 AA; 164887 MW; E6F5710DEDFCB831 CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1477;  
 Best Local Similarity 77.3%; Pred. No. 2.2e-06;  
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VENVYFIRAHDSVQVTRIAKII 22  
 DB 603 IENVYFIRAHDSVQVTRIAKII 624

RESULT 4

Q9EZHS PRELIMINARY; PRT; 1508 AA.  
 AC Q9EZHS; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Dextranucrase Dsrb742.  
 GN Dsrb742.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF294469; P:glucan biosynthesis; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding\_1; 13.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87A6F3A CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1508;  
 Best Local Similarity 77.3%; Pred. No. 2.2e-06;  
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPMVFIRADSEVQVRIAKII 22  
 Db 634 IPNYSFVRADSEVQVIAQII 655

RESULT 5  
 OS2224 PRELIMINARY; PRT; 1508 AA.

ID OS2224  
 AC OS2224;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase (EC 2.4.1.5).  
 GN DSRB.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1299;  
 RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
 RT "Cloning and sequencing of a gene coding for an extracellular  
 RT dextranucrase (DSRB) from leuconostoc mesenteroides NRRL B-1299  
 RT synthesizing only a a(1-6) glucan."  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 DR EMBL: AF030129; AAB95453.1; -.  
 DR PIR: T31098;  
 DR GO: GO:0016757; P:transferase activity, transferring glucosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding\_1; 13.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CEGB57A70D1F0 CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1508;  
 Best Local Similarity 77.3%; Pred. No. 2.2e-06;  
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPMVFIRADSEVQVRIAKII 22  
 Db 634 IPNYSFVRADSEVQVIAQII 655

RESULT 6  
 OS2224 PRELIMINARY; PRT; 2835 AA.  
 ID OS2224  
 AC OS2224;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Dextranucrase (EC 2.4.1.5) (Fragment).  
 GN DSRB.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22231661; PubMed=12270834;  
 RA Bozonnet S., Dolz-lafargue M., Fabre E., Pizzut S., Renaud-Simeon M.,  
 RA Mondan P., Willemot R.M.;  
 RT "Molecular characterization of DSR-E, an alpha-1,2 linkage  
 RT synthesizing dextranucrase with two catalytic domains."  
 RL J. Bacteriol. 184:5753-5761 (2002).  
 DR EMBL: AJ430204; CAD22883.1; -.  
 DR GO: GO:0016757; P:transferase activity, transferring glucosyl. . .; IEA.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding\_1; 20.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 2.  
 KM Transferase; Glycosyltransferase.  
 FT NON\_TER  
 SQ SEQUENCE 2835 AA; 313264 MW; D03262CDD735399D CRC64;

Query Match 81.8%; Score 90; DB 2; Length 2835;  
 Best Local Similarity 85.7%; Pred. No. 4.6e-06;  
 Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 PMVFIRADSEVQVRIAKII 22  
 Db 629 PMVFIRADSEVQVIAQII 649

RESULT 7  
 OS2224 PRELIMINARY; PRT; 1527 AA.  
 ID OS2224  
 AC OS2224;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Dextranucrase.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512-F;  
 RA Bhattachar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
 RT Leuconostoc mesenteroides NRRL B-512F."  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U81374; AAD10952.1; -.  
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.  
 DR GO: GO:0002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW binding\_1; 12.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 80.9%; Score 89; DB 2; Length 1527;  
 Best Local Similarity 72.7%; Pred. No. 3.4e-06;  
 Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PMVFIRADSEVQVRIAKII 22  
 Db 652 PMVFIRADSEVQVIAQII 673

RESULT 8  
 OS2224 PRELIMINARY; PRT; 1527 AA.  
 ID OS2224  
 AC OS2224;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Dextranucrase DsrD (EC 2.4.1.5).  
 GN DSRD.

OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Neubauer H., Bauche A., Mollet B.;  
RT "Isolation and characterization of the dextranucrase Dsd of  
Leuconostoc mesenteroides Lcc4";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A017384; AAG61158.1; -.  
DR GO; GO:001675; P:transferase activity, transferring glycosyl. .; IEA.  
DR GO; GO:000250; P:glucan biosynthesis; IEA.  
DR InterPro: IPR002479; CW binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 12.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Transferrase; Glycosyltransferase.  
SQ SEQUENCE 1527 AA; 169835 MW; F9D0DE220BD89668 CRC64;

Query Match 80.9%; Score 89; DB 2; Length 1527;  
Best Local Similarity 72.7%; Pred. No. 3.4e-06;  
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PNYVFIKRAHDSVQTRIARI 22  
Db 652 PNYSVFRAHDSVQTVIAQIV 673

RESULT 9  
ID 09WKJ5 PRELIMINARY; PRT; 1512 AA.  
AC 09WKJ5;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DE 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE GTF-S.  
GN GTF.  
OS Streptococcus criceti.  
OC Plasmid PAM1.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase (gfts and gtf) genes";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A8026123; BAA77237.1; -.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro: IPR002479; CW binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 11.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid  
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CB601FC14 CRC64;

Query Match 80.0%; Score 88; DB 2; Length 1512;  
Best Local Similarity 81.0%; Pred. No. 5e-06;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PNYVFIKRAHDSVQTRIARI 22  
Db 560 PNYVFIKRAHDSVQTVIAQII 580

RESULT 10  
ID 09LCU7 PRELIMINARY; PRT; 1016 AA.  
AC 09LCU7;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DS Dextranucrase.  
GN DSRT.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRL B-512F;  
RX MEDLINE=20169623; PubMed=10705445;  
RA Fune K., Mizuno K., Takahara H., Kobayashi M.;  
RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
mesenteroides NRL B-512F";  
RL Biosci. Biotechnol. Biochem. 64:29-38 (2000).  
DR EMBL; A8020020; BAA90527.1; -.  
DR HSRP; P06278; IVS.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1016 AA; 110344 MW; 8896FDE13CCCB47 CRC64;

Query Match 79.1%; Score 87; DB 2; Length 1016;  
Best Local Similarity 81.0%; Pred. No. 4.8e-06;  
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PNYVFIKRAHDSVQTRIARI 22  
Db 625 PNYSVFRAHDSVQTVIAQII 645

RESULT 11  
ID 08KZL5 PRELIMINARY; PRT; 1554 AA.  
AC 08KZL5;  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Glucosyltransferase.  
GN GTFU.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21958684; PubMed=11960691;  
RA Hanada N., Fukushima K., Nomura Y., Sempuku H., Hayakawa M.,  
RT "Cloning and nucleotide sequence analysis of the Streptococcus  
sobrinus gtfu gene that produces a highly branched water-soluble  
glucan";  
RL Biochim. Biophys. Acta 1570:75-79 (2002).  
DR EMBL; A8089438; BAC07265.1; -.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR InterPro: IPR002479; CW binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferrase.  
SQ SEQUENCE 1554 AA; 171676 MW; 6981BC1DAE24A73 CRC64;

Query Match 79.1%; Score 87; DB 2; Length 1554;  
Best Local Similarity 72.7%; Pred. No. 7.7e-06;  
Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PNYVFIKRAHDSVQTRIARI 22  
Db 557 PNYSVFRAHDSVQTVIAQIV 578

RESULT 12

Q48756  
ID Q48756 PRELIMINARY; PRT; 1290 AA.  
AC Q48756;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE Dextranucrase.  
OC Bacteroides mesenteroides.  
OC Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B1299;  
RX MEDLINE=97136686; PubMed=8982063;  
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;  
RT "Cloning and sequencing of a gene coding for the alpha dextranucrase  
from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
6) and alpha (1-3) linkages.";  
RL Gene 182:23-32(1996).  
DR EMBL; U38181; AAA40875.1; -.  
DR PIR; JCS473; JCS473.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CM binding.  
DR Pfam; PF01473; CM binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1290 AA; 145590 MM; 3555C2E96B749FAA CRC64;

Query Match 78.2%; Score 86; DB 2; Length 1290;  
Best Local Similarity 85.7%; Pred. No. 9.4e-06;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 NYVFIRAHSEVQTRIAXII 22  
Db 388 NYVFIRAHSEVQTRIAXII 408

RESULT 13  
Q9LCH3 PRELIMINARY; PRT; 1575 AA.  
AC Q9LCH3;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Glucosyltransferase.  
GN GTFR.  
OS Streptococcus oralis.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OC NCBI\_TaxID=1303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC10557;  
RX MEDLINE=20231779; PubMed=10768934;  
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;  
RT "Purification, characterization, and molecular analysis of the gene  
RT encoding glucosyltransferase from Streptococcus oralis.";  
RL Infect. Immun. 68:2475-2483(2000).  
DR EMBL; AB025228; BA95201.1; -.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CM binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CM binding\_1; 16.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1575 AA; 176792 MM; 772A26B4D7C2B543 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1575;  
Best Local Similarity 80.0%; Pred. No. 3.9e-05;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 NYVFIRAHSEVQTRIAXII 22

Db 619 NYVFIRAHSEVQTVIADII 638

RESULT 14  
Q54178 PRELIMINARY; PRT; 1577 AA.  
AC Q54178; Q54247;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Glucosyltransferase.  
GN GTFG.  
OS Streptococcus gordonii Challis.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OC NCBI\_TaxID=29390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=96157084; PubMed=8586195;  
RA Vackerman M.M., Sulavik M.C., Clewell D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
RT phase variants.";  
RL Dev. Biol. Stand. 85:309-314(1995).  
RN [2]  
RP SEQUENCE OF 1-96 FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=92276337; PubMed=1534326;  
RA Sulavik M.C., Tardif G., Clewell D.B.;  
RT "Identification of a gene, 199, which regulates expression of  
RT glucosyltransferase and influences the Spp phenotype of Streptococcus  
RT gordonii Challis.";  
RL J. Bacteriol. 174:3577-3586(1992).  
DR EMBL; U12643; AAC43483.1; -.  
DR EMBL; M89776; AAA26969.1; -.  
DR PIR; B41898; B41898.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CM binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CM binding\_1; 13.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1577 AA; 177805 MM; 5AB0328DC5E08D18 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1577;  
Best Local Similarity 80.0%; Pred. No. 3.9e-05;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 NYVFIRAHSEVQTVIADII 640  
Db 621 NYVFIRAHSEVQTVIADII 640

RESULT 15  
Q55265 PRELIMINARY; PRT; 1577 AA.  
AC Q55265;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Glucosyltransferase precursor.  
GN GTFM.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OC NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=951212197; PubMed=7822030;  
RA Simpson C.L., Giffard P.M., Jacques N.A.;  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes

```

RT coding for primer-independent glucosyltransferases."
RL Infect. Immun. 63:609-621(1995).
DR EMBL; J35928; AAC1413.1; -.
DR PIR; T30858; T30858.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; Ceurface antigen.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR ProDom; PD153432; Ceurface_antigen; 1.
DR Signal; Transferase.
FT CHAIN 1 38 POTENTIAL.
FT SEQUENCE 1577 AA; 175290 MW; 3EFB98A7D3A7B73 CRC64;

Query Match
Best Local Similarity 70.0%; Score 77; DB 2; Length 1577;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFRADSEVQTRIAKII 22
DB 661 NYVFRADSEVQAVLIANII 680

RESULT 16
ID Q00599 PRELIMINARY; PRT; 1599 AA.
AC Q00599;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Glucosyltransferase S precursor (EC 2.4.1.5) (GTF) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFK.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE=93381463; PubMed=8371114;
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and
RT evolution of the gtf genes of oral streptococci."
RT J. Gen. Microbiol. 139:1511-1522(1993).
RN [2]
RP SOURCE FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least two
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975."
RT J. Gen. Microbiol. 137:2577-2593(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO
CC PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF
CC THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL) (N) = D-
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- DISORDER: DENTAL CARIES.
CC -1- SIMILARITY: TO REGIONS OF EARLY AND BACILLUS AMYLOLIQUEFACIENS
CC ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.
DR EMBL; Z11872; CAA77898.1; -.
DR EMBL; Z11873; CAA77901.1; -.
DR EMBL; M6411; AAA26897.1; -.
DR PIR; S22737; S22737.
DR GO; GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.

```

```

DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 13.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 42 POTENTIAL.
FT CHAIN 1 42
FT SEQUENCE 1599 AA; 176480 MW; 24B77869E152B707 CRC64;

Query Match
Best Local Similarity 78.9%; Score 75; DB 2; Length 1599;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 YVFRADSEVQTRIAKII 22
DB 575 YVFRADSEVQTVIADII 593

RESULT 17
ID Q68542 PRELIMINARY; PRT; 1449 AA.
AC Q68542;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Glucosyltransferase N (Fragment).
GN GTFN.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=V1477;
RA Jaffe R.I.;
RT "Streptococcus salivarius V1477 gtfN";
RT Submitted (FEH-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049609; AAC05156.1; -.
DR PIR; T30552; T30552.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
FT NON_TER 1449 1449
FT SEQUENCE 1449 AA; 159895 MW; 0700F6D746471BFB CRC64;

Query Match
Best Local Similarity 70.0%; Score 74; DB 2; Length 1449;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFRADSEVQTRIAKII 22
DB 609 NYVFRADSEVQITIGQII 628

RESULT 18
ID Q55264 PRELIMINARY; PRT; 1449 AA.
AC Q55264;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN GTFK.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN (1)
RP SOURCE FROM N.A.
RX MEDLINE=95122197; PubMed=7822030;

```

RA Simpson C.L., Giffard P.M., Jacques N.A.;  
 "Streptococcus salivarius ATCC 25975 possesses at least two genes  
 coding for primer-independent glucosyltransferases.";  
 RT Infect. Immun. 63:609-621(1995).  
 DR EMBL; J35495; AAC1412.1; -.  
 DR PIR; T30857; T30857.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Signal; Transferase.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.  
 SQ SEQUENCE 1449 AA; 159984 MW; D662F0730686A46 CRC64;  
 Query Match 67.3%; Score 74; DB 2; Length 1449;  
 Best Local Similarity 70.0%; Pred. No. 0.0013;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 NYVFIRAHDSVQTRIAKII 22  
 Db 609 NYAFVRAHDSVQSIIGQII 628  
 RESULT 19  
 ID Q00600 PRELIMINARY; PRT; 1518 AA.  
 AC Q00600;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase I (EC 2.4.1.5) (GTF) (Dextranucrase) (Sucrose 6-  
 DE Glucosyltransferase).  
 DE GTFJ.  
 OS Streptococcus salivarius.  
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 ON NCBI\_TaxID=1304;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25975;  
 RX MEDLINE=92148377; PubMed=1838391;  
 RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;  
 RT "Molecular characterization of a cluster of at least two  
 RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
 RL J. Gen. Microbiol. 137:2577-2593(1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO  
 CC PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF  
 CC THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL) (N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL) (N+1).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- DISEASE: DENTAL CARIES.  
 DR EMBL; Z11873; CAAT7900.1; -.  
 DR EMBL; W6411; AAA26896.1; -.  
 DR PIR; A44811; A44811.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_I; 11.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase; Glucosyltransferase; Repeat; Dental caries.  
 XM DOMAIN 1307 1482 6 DIRECT REPEATS.  
 FT REPEAT 1.307 1338 REPEAT 1.  
 FT REPEAT 1.339 1352 REPEAT 2.  
 FT REPEAT 1.372 1403 REPEAT 3.  
 FT REPEAT 1.404 1417 REPEAT 4.  
 FT REPEAT 1.437 1468 REPEAT 5.  
 FT REPEAT 1.469 1482 REPEAT 6.

SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;  
 Query Match 66.4%; Score 73; DB 2; Length 1518;  
 Best Local Similarity 75.0%; Pred. No. 0.0021;  
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 NYVFIRAHDSVQTRIAKII 22  
 Db 604 NYVFIRAHDSVQDIIRII 623  
 RESULT 20  
 ID Q59983 PRELIMINARY; PRT; 1590 AA.  
 AC Q59983;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5).  
 DE GTFJ.  
 OS Streptococcus sobrinus.  
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 ON NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OMZ176;  
 RX MEDLINE=94146405; PubMed=8312602;  
 RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
 RT "DNA sequence of the glucosyltransferase gene of serotype d  
 RT Streptococcus sobrinus.";  
 RL DNA Seq. 4:19-27(1993).  
 DR EMBL; D13858; BAA02976.1; -.  
 DR PIR; A39841; A39841.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_I; 13.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Glucosyltransferase; Signal; Transferase.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.  
 SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF32B0E CRC64;  
 Query Match 62.7%; Score 69; DB 2; Length 1590;  
 Best Local Similarity 68.2%; Pred. No. 0.011;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 VPMYFIRAHDSVQTRIAKII 22  
 Db 548 VPMYFIRAHDSVQDIIRII 569  
 RESULT 21  
 ID Q55263 PRELIMINARY; PRT; 1590 AA.  
 AC Q55263;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE GTF-I.  
 OS GLUCOSYLTRANSFERASE.  
 OS Streptococcus sobrinus.  
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 ON NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33478;  
 RA Sato S.;  
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase  
 RT produced from Streptococcus sobrinus ATCC 33478.";

Rt	Ann. Kagoshima Univ. Dental School 16:23-29(1996).
DR	EML; D63570; BAA09792.1; -.
DR	PIR; A39841; A39841.
DR	GO; GO:IP090250; P:glucan biosynthesis; IEA.
DR	InterPro; IPR002479; CW_binding.
DR	InterPro; IPR003318; Glyco_hydro_70.
DR	Pfam; PF01473; CW_binding_I; 13.
DR	Pfam; PF02324; Glyco_hydro_70; 1.
SO	SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;
Query Match	62.7%; Score 69; DB 2; Length 1590;
Best Local Similarity	68.2%; Pred.No. 0.011;
Matches 15; Conservative 1; Mismatches	6; Indels 0; Gaps 0
Cy	1 VPNTVFIRAHNSEVQTRIAKII 22    :
Db	548 VPSYSFARAHNSEVDITRII 569

[illegible]

```

RN      SEQUENCE FROM N.A.
RP      [1]
RC      STPAIN-cv. Lemont;
RA      Vysobetskaya V.S., Schwartz J.R., Osborne B.I., Wing R.,
RA      Liu S., Lee Y., Tortum M., Luros J., Li J., Kremetskaya I., Oji O.,
RA      Theologis A.;
RT      "Oryza sativa chromosome 1 BAC 10A191."
RL      Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AC007858; AB039596.1; -.
DR      Gramene; Q9KHV5; -.
DR      InterPro; IPR001623; DnaJ_N.
DR      SMART; SMO0271; DnaJ; 1.
DR      PROSITE; PSS0076; DNaJ 2; 1.
SQ      SEQUENCE      1442 AA; 159434 MW; F63552BDD029454 CRC64;

Query Match      45.5%; Score 50; DB 10; Length 1442;
Beet local Similarity 47.4%; Pred. No. 20;
Matches      9; Conservative      3; Mismatches      7; Indels      0; Gaps      0

Cv      2 PNVFIRAHDEYOTRIAX 20
      |:|:|:|:|:|:|:|:|:|:|
Db      321 PDIAFLKVSQNVQIQTWK 339

```

ID	Q9MXX1	PRELIMINARY;	PR;	492 AA.
AC	Q9MXX1;			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Sugar kinase, FGGY family.			
GN	TM016.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.			
OX	NCBI_Taxid=2336;			
RV	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MSB / DSM 3109;			
RA	MEDLINE=99287316; PubMed=10360571;			
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,			
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,			
RA	McDonald L., Utterback T.R., Malek J.A., Ihnher K.D., Garrett M.M.,			
RA	Stewart A.M., Corton M.D., Pratt M.S., Phillips C.A., Richardson D.,			
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,			
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;			
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from			
RL	genome sequence of Thermotoga maritima."			
DR	Nature 399:323-329(1999).			
DR	EMBL; AE001697; AAD55210.1; -			
DR	PIR; C7417; C7417.			
DR	TIGR; TM016; -			
DR	GO; GO:0003690; P:double-stranded DNA binding; IEA.			
DR	GO; GO:0016301; P:kinase activity; IEA.			
DR	GO; GO:0004856; P:xylokinase activity; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	GO; GO:0006265; P:DNA topological change; IEA.			
DR	GO; GO:0005997; P:xylose metabolism; IEA.			
DR	InterPro; IPR000577; FGGY_kin.			
DR	InterPro; IPR001448; SASP.			
DR	InterPro; IPR006000; Xylulokinase.			
DR	Pfam; PF00370; FGGY_1.			
DR	Pfam; PF02782; FGGY_C_1.			
DR	TIGRFAMS; TIGR01312; xylB_1.			
DR	PROSITE; PS00933; FGGY_KINASES_1; 1.			
DR	PROSITE; PS00445; FGGY_KINASES_2; 1.			
DR	PROSITE; PS00304; SASP_1; 1.			
KM	Kinase; Complete proteome.			
SQ	SEQUENCE 492 AA; 54905 MW; 0F66A3AB451D08E1 CRC64;			
Query Match	43.6%; Score 48; DB 16; Length 492;			
Best Local Similarity	36.4%; Pred. No. 14;			
Matches	8; Conservative			
	8; Mismatches			
	6; Indels			
	0; Gaps			
	0;			

QY 1 VENTPIRAHDSVOTRIAKI 22  
 DB 136 LPKILWIRKHEPEYKISKIM 157

RESULT 25

Q81MG5 PRELIMINARY; PRT; 279 AA.

AC Q81MG5; 081MG5; 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE CG7948-BB.

GN RAD51 OR CG7948.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Phylloidea; Drosophilidae; Drosophila.

OX NCBI\_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazee R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gaber G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelle S., Fleischmann W.,

RA Fostel C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Idbegyan C.,

RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupel M.P., Smith T.,

RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach U.,

RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Carlson J., An H., Balwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.M., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreesnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Idbegyan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Paclet J., Pargas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.M., Celinker S.E.,

RA Clamp M., Drysdale R., Emsert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith B., Shu S., Smutnick F., Whitfield E., S.E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.D., Lewis S.E.;

RT "Annotation of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA FlyBase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL: AAC03772; FAN14213.1; -.

DR FlyBase; Fggn0011700; Rad51.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0003684; P:damaged DNA binding; IEA.

DR GO; GO:0008094; P:DNA dependent ATPase activity; IEA.

DR GO; GO:0000166; F:nucleotide binding; IEA.

DR GO; GO:0006310; P:DNA recombination; IEA.

DR GO; GO:0006281; P:DNA repair; IEA.

DR InterPro; IPR003593; AAA ATPase.

DR SMART; SM00382; AAA; 1.

DR PROSITE; PS50162; RECA\_2; 1.

DR PROSITE; PS50163; RECA\_3; 1.

SQ SEQUENCE 279 AA; 30459 MW; BF82A5C319AB61F9 CRC64;

Query Match 42.7%; Score 47; DB 5; Length 279;

Best Local Similarity 42.1%; Pred. No. 11;

Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 NYVEPIRAHDSVOTRIAKI 21

DB 128 NVAFTRAHNSDOTKILQM 146

RESULT 26

Q81IG58 PRELIMINARY; PRT; 284 AA.

AC Q81IG58; 081IG58; 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE RH24133P.

GN RAD51 OR CG7948.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Phylloidea; Drosophilidae; Drosophila.

OX NCBI\_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Y;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Dreesnek D., Farfan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,

RA Miranda A., Mungall C.J., Pacleb J., Pargas V., Park S.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,

RA Celinker S.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: BF001791; AAN71546.1; -.



DR FlyBase: FBgn0011700; Rad51.  
DR GO: GO:0005524; F:ATP binding; IEA.  
DR GO: GO:0003684; F:damaged DNA binding; IEA.  
DR GO: GO:0008094; F:DNA dependent ATPase activity; IEA.  
DR GO: GO:0000166; F:nucleotide binding; IEA.  
DR GO: GO:0006310; P:DNA recombination; IEA.  
DR GO: GO:0006281; P:DNA repair; IEA.  
DR InterPro: IPR003593; AAA ATPase.  
DR InterPro: IPR001553; RecA.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS50162; RECA\_2; 1.  
DR PROSITE: PS50163; RECA\_3; 1.  
SQ SEQUENCE 284 AA; 30909 MW; 90E63CD591B7FDA9 CRC64;

Query Match 42.7%; Score 47; DB 5; Length 284;  
Best Local Similarity 42.1%; Pred. No. 11;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 NYVFIRAH-DESVOTRIAKT 21  
DB 128 NYVFIRAH-DESVOTRIAKT 146

RESULT 27  
ID 08SX46 PRELIMINARY; PRT; 783 AA.  
AC 08SX46;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE RB09708P.  
OS Drosophila melanogaster (Fruit fly).  
GN CG10051.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,  
RA Champagne M., Chavez C., Dorsett V., Drenth D., Farfan D., Fries E.,  
RA George R., Gonzalez M., Guatin H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Munro K., Pacle J., Paragas V., Park S.,  
RA Patel S., Pounavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY094849; J011202.1; -  
DR FlyBase: FBgn0034437; CG10051.  
DR GO: GO:0008233; P:peptidase activity; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR007484; Peptidase M28.  
DR Pfam: PF04389; Peptidase M28; 1.  
SQ SEQUENCE 783 AA; 89225 MW; E8258160A3674494 CRC64;

Query Match 41.4%; Score 45.5; DB 5; Length 783;  
Best Local Similarity 55.6%; Pred. No. 63;  
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 3 NYVFIRAH-DESVOTRIAKT 19  
DB 72 NYLVNSHYDSVOTPIAA 89

RESULT 28  
ID 09V8T7 PRELIMINARY; PRT; 866 AA.  
AC 09V8T7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CG10051 protein.  
GN CG10051.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RL MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Bencs P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadelin E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck U.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhou Q., Zheng L.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RL "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
DR EMBL: AE003796; AA57573.1; -  
DR FlyBase: FBgn0034437; CG10051.  
DR GO: GO:0008233; P:peptidase activity; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR007484; Peptidase M28.  
DR Pfam: PF04389; Peptidase M28; 1.  
SQ SEQUENCE 866 AA; 98558 MW; 19C76A6569B9141A CRC64;

Query Match 41.4%; Score 45.5; DB 5; Length 866;  
Best Local Similarity 55.6%; Pred. No. 70;  
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 3 NYVFIRAH-DESVOTRIAKT 19  
DB 155 NYLVNSHYDSVOTPIAA 172

RESULT 29  
ID 09A9H5 PRELIMINARY; PRT; 498 AA.  
AC 09A9H5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein CC1005 (Sat1).  
GN CC1005 OR SST1.

OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,  
Petrocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
Ruttenberg T., Tran K., Wolf A., Vamathevan J., Ermolova M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,  
"Complete genome sequence of Caulobacter crescentus.";  
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15, and CB15N / NA1000;  
RX MEDLINE=98292737; PubMed=9620954;  
RA Awram P., Smit J.;  
RT "The Caulobacter crescentus paracrystalline S-layer protein is  
secreted by an ABC transporter (type I) secretion apparatus.";  
J. Bacteriol. 180:3062-3069(1998).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15, and CB15N / NA1000;  
RX MEDLINE=21286856; PubMed=11390676;  
RA Awram P., Smit J.;  
RT "Identification of lipopolysaccharide O antigen synthesis genes  
required for attachment of the S-layer of Caulobacter crescentus.";  
Microbiology 147:1451-1460(2001).  
DR EMBL; AE005778; AAK2989.1; -.  
DR EMBL; AF062345; AAK2615.1; -.  
DR PIR; A87374; A87374.  
DR TIGR; CC1005; -.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 498 AA; 51975 MW; 42B55228D851A67 CRC64;  
Query Match 40.9%; Score 45; DB 16; Length 498;  
Best Local Similarity 50.0%; Pred. No. 46;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
CY 2 PNYVYFIRAHDSVQTRIAKI 21  
Db 56 PMFVFYRADGDEVYIHFAFI 75  
RESULT 30  
Q9AQR3 PRELIMINARY; PRT; 639 AA.  
AC Q9AQR3;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Protease.  
GN PROA.  
OS Bacillus sp. 9860.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=133778;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9860;  
RX MEDLINE=20568675; PubMed=1118284;  
RA Saeki K., Okada M., Hatada Y., Kobayashi T., Ito S., Takami H.,  
Horikoshi K.;  
RT "Novel oxidatively stable subtilisin-like serine proteases from  
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and  
evolutionary relationships.";  
Biochem. Biophys. Res. Commun. 279:313-319(2000).  
DR EMBL; AB046403; BAB21266.2; -.  
DR HSSP; P00782; 1SUF.

DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000208; Peptidase\_S8.  
DR InterPro; IPR007280; PPC.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
SQ SEQUENCE 639 AA; 68185 MW; 316AF6FDBE4FF54 CRC64;  
Query Match 40.9%; Score 45; DB 2; Length 639;  
Best Local Similarity 28.6%; Pred. No. 61;  
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;  
CY 1 VNYVYFIRAHDSVQTRIAKI 21  
Db 114 IPDYAYIVEYEGDVQSKVRSI 134  
Search completed: May 4, 2004, 09:12:36  
Job time : 37 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:07:31 ; Search time 14 Seconds  
(without alignments)

81.127 Million cell updates/sec

Title: US-09-290-049A-18  
Perfect score: 110  
Sequence: 1 VPNYVFIKRAHSEVQTRIAKTI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/iaa/PCTus\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	90	81.8	545 4	US-09-604-957-4 Sequence 4, Appli
2	90	81.8	1430 3	US-09-008-172-2 Sequence 2, Appli
3	90	81.8	1430 3	US-09-210-361-6 Sequence 6, Appli
4	90	81.8	1430 4	US-09-740-274-6 Sequence 6, Appli
5	89	80.9	523 4	US-09-604-957-5 Sequence 5, Appli
6	80	72.7	1475 3	US-09-007-999-2 Sequence 2, Appli
7	80	72.7	1475 3	US-09-210-361-2 Sequence 2, Appli
8	80	72.7	1475 4	US-09-740-274-2 Sequence 2, Appli
9	77	70.0	1577 2	US-08-793-824-2 Sequence 2, Appli
10	76	69.1	1375 3	US-09-210-361-4 Sequence 4, Appli
11	76	69.1	1375 4	US-09-740-274-4 Sequence 4, Appli
12	65	59.1	584 4	US-09-604-957-6 Sequence 7, Appli
13	63	57.3	535 4	US-09-604-957-7 Sequence 6, Appli
14	63	57.3	1278 4	US-09-604-957-3 Sequence 3, Appli
15	63	57.3	2057 4	US-09-499-203-2 Sequence 2, Appli
16	45	40.9	639 4	US-09-509-814A-4 Sequence 4, Appli
17	42	38.2	183 4	US-09-107-532A-5511 Sequence 5511, Ap
18	42	38.2	652 4	US-09-328-352-5587 Sequence 5587, Ap
19	42	38.2	3118 4	US-09-579-181-1 Sequence 1, Appli
20	41	37.3	416 3	US-09-330-611-2 Sequence 1, Appli
21	41	37.3	345 3	US-09-489-039A-11050 Sequence 11050, A
22	40	36.4	48 3	US-09-330-611-21 Sequence 21, Appli
23	40	36.4	373 4	US-09-489-039A-9606 Sequence 9606, Ap
24	40	36.4	943 4	US-08-911-321-4 Sequence 4, Appli
25	40	36.4	1041 1	US-08-220-151-4 Sequence 4, Appli
26	40	36.4	1041 1	US-08-413-118-4 Sequence 4, Appli
27	40	36.4	1041 3	US-08-473-446-4 Sequence 4, Appli

28	40	36.4	1231 4	US-09-071-035-420 Sequence 420, App
29	40	36.4	1265 4	US-09-071-035-418 Sequence 418, App
30	40	36.4	1278 4	US-09-134-000C-6043 Sequence 6043, Ap
31	39	35.5	85 4	US-09-489-039A-10868 Sequence 10868, A
32	39	35.5	195 4	US-09-328-352-4777 Sequence 4777, Ap
33	39	35.5	287 2	US-08-901-306-2 Sequence 2, Appli
34	39	35.5	287 3	US-09-180-271-2 Sequence 2, Appli
35	39	35.5	533 4	US-09-134-001C-3592 Sequence 3592, Ap
36	39	35.5	760 4	US-09-489-039A-8055 Sequence 8055, Ap
37	39	35.5	1127 4	US-09-252-991A-26849 Sequence 26849, A
38	39	34.5	75 4	US-09-107-532A-7020 Sequence 7020, Ap
39	38	34.5	89 4	US-09-489-039A-12442 Sequence 12442, A
40	38	34.5	144 4	US-09-621-976-7647 Sequence 7647, Ap
41	38	34.5	186 4	US-09-252-991A-21822 Sequence 21822, A
42	38	34.5	233 4	US-09-252-991A-24679 Sequence 24679, A
43	38	34.5	368 3	US-08-709-838-2 Sequence 2, Appli
44	38	34.5	368 3	US-08-829-839-2 Sequence 20, Appli
45	38	34.5	368 4	US-09-170-496D-20 Sequence 174, App
46	38	34.5	368 4	US-09-170-496D-174 Sequence 2, Appli
47	38	34.5	368 4	US-09-624-594-2 Sequence 4386, Ap
48	38	34.5	416 4	US-09-543-681A-4386 Sequence 17, Appli
49	38	34.5	454 4	US-09-198-452A-197 Sequence 17, Appli
50	38	34.5	615 2	US-08-663-566A-17 Sequence 17, Appli
51	38	34.5	615 2	US-08-023-610-17 Sequence 17, Appli
52	38	34.5	615 2	US-08-288-065A-17 Sequence 17, Appli
53	38	34.5	615 2	US-08-362-240A-17 Sequence 17, Appli
54	38	34.5	615 5	PCT-US95-10245-17 Sequence 17, Appli
55	38	34.5	670 4	US-09-328-352-6725 Sequence 2548, A
56	38	34.5	679 4	US-09-252-991A-29548 Sequence 2, Appli
57	38	34.5	868 2	US-08-907-166-2 Sequence 2, Appli
58	38	34.5	868 4	US-09-391-340-2 Sequence 11, Appli
59	38	34.5	1162 2	US-08-663-566A-15 Sequence 15, Appli
60	38	34.5	1162 2	US-08-023-610-15 Sequence 15, Appli
61	38	34.5	1162 2	US-08-288-065A-15 Sequence 15, Appli
62	38	34.5	1162 2	US-08-362-240A-15 Sequence 15, Appli
63	38	34.5	1162 5	PCT-US95-10245-15 Sequence 15, Appli
64	38	34.5	1464 3	US-08-891-640-2 Sequence 33113, A
65	37.5	34.1	332 4	US-09-252-991A-33113 Sequence 6201, Ap
66	37.5	34.1	674 4	US-09-107-532A-6201 Sequence 7370, Ap
67	37	33.6	62 4	US-09-489-039A-7370 Sequence 682, App
68	37	33.6	112 4	US-09-198-452A-682 Sequence 7342, App
69	37	33.6	145 4	US-09-328-352-7342 Sequence 8261, Ap
70	37	33.6	254 4	US-09-543-681A-8261 Sequence 1126, Ap
71	37	33.6	256 4	US-09-198-452A-1126 Sequence 7011, Ap
72	37	33.6	294 4	US-09-107-532A-7011 Sequence 4336, Ap
73	37	33.6	352 4	US-09-107-532A-4936 Sequence 7543, Ap
74	37	33.6	359 4	US-09-328-352-7543 Sequence 1604, Ap
75	37	33.6	461 4	US-09-134-001C-3604 Sequence 32, Appli
76	37	33.6	545 4	US-09-934-901C-12 Sequence 2, Appli
77	37	33.6	553 1	US-08-475-894-2 Sequence 2, Appli
78	37	33.6	553 1	US-08-484-710-2 Sequence 2, Appli
79	37	33.6	553 2	US-08-484-709-2 Sequence 2, Appli
80	37	33.6	553 3	US-08-474-697-2 Sequence 2, Appli
81	37	33.6	559 4	US-09-134-001C-5572 Sequence 5572, Ap
82	37	33.6	588 4	US-09-650-324A-58 Sequence 38, Appli
83	37	33.6	616 4	US-09-268-347-38 Sequence 46, Appli
84	37	33.6	659 4	US-09-268-347-46 Sequence 46, Appli
85	37	33.6	660 4	US-09-268-347-45 Sequence 1692, A
86	37	33.6	903 4	US-09-489-039A-13692 Sequence 2, Appli
87	37	33.6	936 5	PCT-US93-0594A-2 Sequence 2, Appli
88	37	33.6	1008 4	US-09-308-453-2 Sequence 3, Appli
89	37	33.6	1118 3	US-09-379-523-3 Sequence 17, Appli
90	37	33.6	1715 4	US-08-626-115B-17 Sequence 35, Appli
91	36.5	33.2	188 1	US-08-442-063A-39 Sequence 42, Appli
92	36.5	33.2	236 1	US-08-442-063A-42 Sequence 5, Appli
93	36.5	33.2	252 4	US-09-395-689-5 Sequence 6897, Ap
94	36.5	33.2	279 4	US-09-107-532A-6807 Sequence 48, Appli
95	36.5	33.2	282 1	US-08-442-063A-45 Sequence 27, Appli
96	36.5	33.2	287 1	US-08-442-063A-48 Sequence 27, Appli
97	36.5	33.2	333 1	US-08-442-063A-27 Sequence 2, Appli
98	36.5	33.2	342 1	US-08-212-919-2 Sequence 2, Appli
99	36.5	33.2	342 1	US-08-619-916-2 Sequence 2, Appli
100	36.5	33.2	342 5	PCT-US95-08542-2 Sequence 2, Appli

## ALIGNMENTS

```
RESULT 1
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERTRUDINA HENDRIKA
; APPLICANT: DIJKUIZEN, LUBBERT
; APPLICANT: RAHOUJI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRF
; ORGANISM: Streptococcus mutans
US-09-604-957-4
```

```
Query Match      81.8%; Score 90; DB 4; Length 545;
Best Local Similarity 90.0%; Pred. No. 1e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 NYVFIRAHDSVQTVIAKII 22
      ||:|||||
DB      156 NYVFIRAHDSVQTVIAKII 175
```

```
RESULT 2
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRF
; ORGANISM: Streptococcus mutans
US-09-008-172-2
```

```
Query Match      81.8%; Score 90; DB 3; Length 1430;
Best Local Similarity 90.0%; Pred. No. 3.2e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 NYVFIRAHDSVQTVIAKII 22
      ||:|||||
DB      576 NYVFIRAHDSVQTVIAKII 595
```

```
RESULT 3
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
```

```
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRF
; ORGANISM: streptococcus mutans
US-09-210-361-6
```

```
Query Match      81.8%; Score 90; DB 3; Length 1430;
Best Local Similarity 90.0%; Pred. No. 3.2e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 NYVFIRAHDSVQTVIAKII 22
      ||:|||||
DB      576 NYVFIRAHDSVQTVIAKII 595
```

```
RESULT 4
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CND
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRF
; ORGANISM: streptococcus mutans
US-09-740-274-6
```

```
Query Match      81.8%; Score 90; DB 4; Length 1430;
Best Local Similarity 90.0%; Pred. No. 3.2e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 NYVFIRAHDSVQTVIAKII 22
      ||:|||||
DB      576 NYVFIRAHDSVQTVIAKII 595
```

```

RESULT 5
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHOTTEN, GERRITJINA HENDRIKA
; APPLICANT: DIKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match      80.9%; Score 89; DB 4; Length 523;
Best Local Similarity 72.7%; Pred. No. 1.5e-07;
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 VPNTVFIRADSEVQTRIAKII 22
Db      146 IPNYSFVRADSEVQTVIAQIV 167

RESULT 6
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match      72.7%; Score 80; DB 3; Length 1475;
Best Local Similarity 77.3%; Pred. No. 2e-05;
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1 VPNTVFIRADSEVQTRIAKII 22
Db      552 VPSYSFIRADSEVQDLIADI 573

RESULT 7
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR

```

```

; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match      72.7%; Score 80; DB 3; Length 1475;
Best Local Similarity 77.3%; Pred. No. 2e-05;
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1 VPNTVFIRADSEVQTRIAKII 22
Db      552 VPSYSFIRADSEVQDLIADI 573

RESULT 8
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match      72.7%; Score 80; DB 4; Length 1475;
Best Local Similarity 77.3%; Pred. No. 2e-05;
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1 VPNTVFIRADSEVQTRIAKII 22
Db      552 VPSYSFIRADSEVQDLIADI 573

RESULT 9

```

US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Giffard, Philip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of Plants to  
; TITLE OF INVENTION: Increase Stored Carbohydrates  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838 Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ. ID NO. 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
; US-08-793-824-2  
  
Query Match 70.0%; Score 77; DB 2; Length 1577;  
Best Local Similarity 70.0%; Pred. No. 7.5e-05;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 NYVYFIRAHDSVQTRIAKII 22  
||:|||||||:|  
Db 661 NYVYFIRAHDSVQAVLANII 680  
  
RESULT 10  
US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; EARLIER FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172

; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
; US-09-210-361-4  
  
Query Match 69.1%; Score 76; DB 3; Length 1375;  
Best Local Similarity 72.7%; Pred. No. 9.6e-05;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 VNNYFIRAHDSVQTRIAKII 22  
||:|||||||:|  
Db 578 VPSYFIRAHDSVQDLIRNII 599  
  
RESULT 11  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. 6465203  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
; US-09-740-274-4  
  
Query Match 69.1%; Score 76; DB 4; Length 1375;  
Best Local Similarity 72.7%; Pred. No. 9.6e-05;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 VNNYFIRAHDSVQTRIAKII 22  
||:|||||||:|  
Db 578 VPSYFIRAHDSVQDLIRNII 599  
  
RESULT 12  
US-09-604-957-6  
; Sequence 6, Application US/09604957  
; Patent No. 6486314  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDIEN HENDRIKA  
; APPLICANT: DIJKUITZEN, LOBBERT  
; APPLICANT: RAHOU, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957

CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 59.1%; Score 65; DB 4; Length 584;  
Best Local Similarity 55.0%; Pred. No. 0.0032;  
Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPNTVFIRADSEVQTRIAK 20  
DB 167 IPNTSFVRAHDYDQADPIRK 186

RESULT 13  
US-09-604-957-7  
Sequence 7, Application US/09604957  
Patent No. 6486314  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOUBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 7  
LENGTH: 535  
TYPE: PRT  
ORGANISM: Lactobacillus reuteri  
US-09-604-957-7

Query Match 57.3%; Score 63; DB 4; Length 535;  
Best Local Similarity 50.0%; Pred. No. 0.0065;  
Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPNTVFIRADSEVQTRIAK 22  
DB 144 IPNTSFVRAHDNNSQDQIQNAI 165

RESULT 14  
US-09-604-957-3  
Sequence 3, Application US/09604957  
Patent No. 6486314  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOUBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1278  
TYPE: PRT  
ORGANISM: Lactobacillus reuteri

US-09-604-957-3

Query Match 57.3%; Score 63; DB 4; Length 1278;  
Best Local Similarity 50.0%; Pred. No. 0.018;  
Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPNTVFIRADSEVQTRIAK 22  
DB 620 IPNTSFVRAHDNNSQDQIQNAI 641

RESULT 15  
US-09-499-203-2  
Sequence 2, Application US/09499203  
Patent No. 6570065  
GENERAL INFORMATION:  
APPLICANT: KOSSMANN, Jens  
APPLICANT: WELSH, Thomas  
APPLICANT: QUANZ, Martin  
APPLICANT: KNUTH, Katola  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternanucrase  
FILE REFERENCE: 147-196P  
CURRENT APPLICATION NUMBER: US/09/499,203  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 57.3%; Score 63; DB 4; Length 2057;  
Best Local Similarity 55.0%; Pred. No. 0.032;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNTVFIRADSEVQTRIAK 20  
DB 757 IPNTSFVRAHDYDQADPIRK 776

RESULT 16  
US-09-509-814A-4  
Sequence 4, Application US/09509814A  
Patent No. 6376227  
GENERAL INFORMATION:  
APPLICANT: TAKAIWA, MIKIO  
APPLICANT: OKUDA, MITSUYOSHI  
APPLICANT: SAKKI, KATSUHIISA  
APPLICANT: KUBOTA, HIROMI  
APPLICANT: HITOMI, JUN  
APPLICANT: KAGEYAMA, YASUSHI  
APPLICANT: SHIKATA, SHITSUMI  
APPLICANT: NOMURA, MASAFUMI  
TITLE OF INVENTION: ALKALINE PROTEASE  
FILE REFERENCE: 0327-0832-0PCT  
CURRENT APPLICATION NUMBER: US/09/509,814A  
CURRENT FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: PCT/JP98/04528  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: JP 9-274570  
PRIOR FILING DATE: 1997-06-08  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 4  
LENGTH: 639  
TYPE: PRT  
ORGANISM: Bacillus sp.  
US-09-509-814A-4

Query Match 40.9%; Score 45; DB 4; Length 639;  
Best Local Similarity 28.6%; Pred. No. 13;  
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 VNVYVFIHDSVQTRIAKI 21  
Db 114 IPDYAVIVEYBGDVDSKVRSI 134

RESULT 17  
US-09-107-532A-5511  
Sequence 5511, Application US/09107532A  
Patent No. 6583275

## GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSER: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Maltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: &lt;Unknown&gt;

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5511:

SEQUENCE CHARACTERISTICS:

LENGTH: 183 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc.feature

LOCATION: (B) LOCATION 1...183

SEQUENCE DESCRIPTION: SEQ ID NO: 5511:

US-09-107-532A-5511

Query Match 38.2%; Score 42; DB 4; Length 183;

Best Local Similarity 41.2%; Pred. No. 10;

Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 5 VFIHDSVQTRIAKI 21

Db 40 IFWEAHHSQYITLAKI 56

RESULT 18  
US-09-328-352-5587  
Sequence 5587, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03BA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 5587

LENGTH: 652

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-5587

Query Match 38.2%; Score 42; DB 4; Length 652;

Best Local Similarity 35.0%; Pred. No. 44;

Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 NVYVFIHDSVQTRIAKI 22  
Db 227 NVYVFIHDSVQTRIAKI 246RESULT 19  
US-09-579-181-1

Sequence 1, Application US/09579181

Patent No. 6365372

GENERAL INFORMATION:

APPLICANT: Chiviva, John

APPLICANT: Yaciuk, Peter

TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)

FILE REFERENCE: 16153-4247

CURRENT APPLICATION NUMBER: US/09/579,181

CURRENT FILING DATE: 2000-05-25

PRIORITY APPLICATION NUMBER: 60/136,620

PRIOR FILING DATE: 1999-05-27

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 1

LENGTH: 3118

TYPE: PRT

ORGANISM: Human

US-09-579-181-1

Query Match 38.2%; Score 42; DB 4; Length 3118;

Best Local Similarity 50.0%; Pred. No. 2.8e+02;

Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 10 HDSEVQTRIAKI 21

Db 51 HEALIEIRIAEL 62

RESULT 20  
US-09-330-611-2  
Sequence 2, Application US/09330611  
Patent No. 6248874  
GENERAL INFORMATION:

APPLICANT: PREY, Perry A.

APPLICANT: RUZICKA, Frank J.

TITLE OF INVENTION: DNA MOLECULES ENCODING BACTERIAL LYSINE 2,3-AMINOMUTASE

FILE REFERENCE: 032026/0476

CURRENT APPLICATION NUMBER: US/09/330,611

CURRENT FILING DATE: 1999-06-11

EARLIER APPLICATION NUMBER: US 09/198,942

PRIOR FILING DATE: 1998-11-24

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 2

LENGTH: 416

TYPE: PRT

ORGANISM: Clostridium subterminale

US-09-330-611-2

Query Match 37.3%; Score 41; DB 3; Length 416;

Best Local Similarity 31.8%; Pred. No. 39;



Matches 7; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
Qy 1 VPMVFIKADSEVQTRIAKI 22  
Db 341 MPNVISQSHDKVILNPFEGVI 362

RESULT 21  
US-09-489-039A-11050

; Sequence 11050, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 11050

; LENGTH: 545

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-11050

Query Match 37.3%; Score 41; DB 4; Length 545;  
Best Local Similarity 50.0%; Pred. No. 54;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 7 IRAHDSVQTRIAKI 22  
Db 204 IASHDRILTRPRIL 219

RESULT 22

US-09-330-611-21

; Sequence 21, Application US/09330611

; Patent No. 6248874

; GENERAL INFORMATION:

; APPLICANT: Frey, Perry A.

; APPLICANT: Ruzicka, Frank J.

; TITLE OF INVENTION: DNA MOLECULES ENCODING BACTERIAL LYSINE 2,3-AMINOMUTASE

; FILE REFERENCE: 032026/0476

; CURRENT APPLICATION NUMBER: US/09/330,611

; CURRENT FILING DATE: 1999-06-11

; EARLIER APPLICATION NUMBER: US 09/198,942

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 21

; LENGTH: 48

; TYPE: PRT

; ORGANISM: Clostridium subterminale

US-09-330-611-21

Query Match 36.4%; Score 40; DB 3; Length 48;  
Best Local Similarity 33.3%; Pred. No. 4.7;  
Matches 7; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 2 PNVYFIKADSEVQTRIAKI 22  
Db 1 PNVVISQSHDKVILNPFEGVI 21

RESULT 23

US-09-489-039A-9606

; Sequence 9606, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 9606

; LENGTH: 373

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-9606

Query Match 36.4%; Score 40; DB 4; Length 373;  
Best Local Similarity 58.3%; Pred. No. 52;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 IRAHDSVQTRIAKI 18  
Db 247 LRTHSSEVQTRV 258

RESULT 24

US-08-911-321-4

; Sequence 4, Application US/08911321

; Patent No. 6010703

; GENERAL INFORMATION:

; APPLICANT: Roger K. Maes and Stephen J. Spatz

; TITLE OF INVENTION: Recombinant Poxvirus

; TITLE OF INVENTION: Vaccine Against

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Ian C. McLeod

; STREET: 2190 Commons Parkway

; CITY: Okemos

; STATE: Michigan

; COUNTRY: USA

; ZIP: 48864

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB

; MEDIUM TYPE: storage

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/911,321

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/096,183

; FILING DATE: July 26, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Ian C. McLeod

; REGISTRATION NUMBER: 20,931

; REFERENCE/DOCKET NUMBER: MSU 4.1-166

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (517) 347-4100

; TELEFAX: (517) 347-4103

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 943

; TYPE: Amino Acid

; STRANDEDNESS: Single

; TOPOLOGY: Linear

; MOLECULE TYPE: Polypeptide

; DESCRIPTION: No

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Feline herpesvirus-1

; STRAIN: 1

? INDIVIDUAL ISOLATE: C-27  
? CELL TYPE: N/A  
?  
? FEATURE:  
? NAME/KEY:  
? LOCATION:  
? IDENTIFICATION METHOD:  
? OTHER INFORMATION: gb  
?  
US-08-911-321-4

Deduced sequence

```
Query Match      36.4%; Score 40; DB 3; Length 943;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches      8; Conservative 5; Mismatches 3; Indels 0; Gaps 0
```

```
QY      4 YVFIRAHDSVQTRIA 19
         | :|:| :|:|
Db      545 YDYIQAHVNEMLSRIA 560
```

```

1      RESULT 25
2      US-08-220-151-4
3      ; Sequence 4, Application US/08220151
4      Patent No. 5529780
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Paoletti, Enzo
7      ; APPLICANT: Limbach, Keith J.
8      ; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
9      ; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR
10     ; NUMBER OF SEQUENCES: 91
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Curtiss, Morris & Safford
13     ; STREET: 530 Fifth Avenue
14     ; CITY: New York
15     ; STATE: NY
16     ; COUNTRY: USA
17     ZIP: 10036
18     ;
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/220,151
26     ; FILING DATE: 30-MAR-1994
27     ; CLASSIFICATION: 435
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: Frommer, William S.
30     ; REGISTRATION NUMBER: 25,506
31     ; REFERENCE/DOCKET NUMBER: 454310-2540
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (212) 840-3333
34     ; TELEFAX: (212) 840-0712
35     ; TELEX: 425066 CURTMS
36     ; INFORMATION FOR SEQ ID NO. 4:
37     ; SEQUENCE CHARACTERISTICS:
38     ; LENGTH: 1041 amino acids
39     ; TYPE: amino acid
40     ; STRANDEDNESS: single
41     ; TOPOLOGY: linear
42     ; MOLECULE TYPE: peptide
43     ; FRAGMENT TYPE: N-terminal
44     ;
45     US-08-220-151-4

```

Query Match	36.4%	Score 40;	DB 1;	Length 1041;
Best Local Similarity	50.0%;	Pred. No. 1.7e+02;		
Matches	8;	Conservative	5;	Mismatches 3;
			Indels	0;
			Gaps	0

QY 4 YVFIRAH DSEVQTRIA 19  
| : | : | : | : | : |  
Db 641 YDYIAHVNMELSRJA 656

RESULT 26  
US-08-413-118-4

Sequence 4 Application US/08413118  
Patent No. 5688920  
GENERAL INFORMATION:  
APPLICANT: PROLETTI, ENZO  
APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/413,118  
FILING DATE: 29-MAR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/220,151  
FILING DATE: 30-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-0712  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1041 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-413-118-4

Query Match	36.4%	Score 40;	DB 1;	Length 1041;
Best Local Similarity	50.0%	Pred. No. 1.7e+02;		
Matches	8;	Conservative	5;	Mismatches 3;
			Indels	0;
			Gaps	0;

QY 4 YVFI RAHDSEVO TRIA 19  
| : | : | : | : | : |  
Db 641 YDYIQA HVNEMLSRIA 656

RESULT 27  
 US-08-473-446-4  
 ; Sequence 4, Application US/08473446  
 ; Patent No. 6017542  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PAOLETTI, ENZO  
 ; APPLICANT: LIMBACH, KEITH J.  
 ; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
 ; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR  
 ; NUMBER OF SEQUENCES: 128  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P. C.  
 ; STREET: 530 FIFTH AVENUE, 25TH FLOOR  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/413,118
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FROMMER, WILLIAM S.
; REGISTRATION NUMBER: 25,506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-473-446-4

Query Match          36.4%; Score 40; DB 3; Length 1041;
Best Local Similarity 50.0%; Pred. No. 1,7e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY      4 YVYVFIKADSEVQTRIA 19
Db      641 YDYIQAHVNMELSRIA 656

RESULT 28
; US-09-071-035-420
; Sequence 420, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 420:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-420

Query Match          36.4%; Score 40; DB 4; Length 1231;
Best Local Similarity 56.2%; Pred. No. 2.1e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 VYVYVFIKADSEVQTR 16
Db      1055 VDNFVRIRPDQEVVT 1070

RESULT 29
; US-09-071-035-418
; Sequence 418, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 418:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-418

Query Match          36.4%; Score 40; DB 4; Length 1265;
Best Local Similarity 56.2%; Pred. No. 2.2e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 VYVYVFIKADSEVQTR 16
Db      1087 VDNFVRIRPDQEVVT 1102

RESULT 30
; US-09-134-000C-6043
; Sequence 6043, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
```

```

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6043
; LENGTH: 1278
; TYPE: PR1
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6043

```

```

Query Match          36.4%; Score 40; DB 4; Length 1278;
Best Local Similarity 56.2%; Pred. No. 2.2e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 VPNTVFIRAHDSVQCT 16
DB 1100 VDNFVARIRPNDQEVVT 1115

```

Search completed: May 4, 2004, 09:14:12  
Job time : 15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using SW model

Run on: May 4, 2004, 09:12:41 ; Search time 35.333 Seconds  
(without alignments)  
172.590 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VNNYFIRAHSEVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

## Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	90	81.8	545 9	US-09-995-749A-10
2	90	81.8	1430 9	US-09-740-274-6
3	89	80.9	522 9	US-09-995-749A-11
4	80	72.7	1475 9	US-09-740-274-2
5	76	69.1	1375 9	US-09-740-274-4
6	63	57.3	535 9	US-09-995-749A-13
7	63	57.3	584 9	US-09-995-749A-12
8	63	57.3	1781 9	US-09-995-749A-2
9	63	57.3	2057 15	US-10-417-280A-2
10	49	44.5	602 12	US-10-282-122A-76653
11	48	43.6	146 12	US-10-425-114-44862
12	48	43.6	170 12	US-10-425-114-61839
13	48	43.6	251 12	US-10-424-599-279743
14	48	43.6	492 15	US-10-369-493-2894
15	47	42.7	506 12	US-10-282-122A-47139

16	47	42.7	557 12	US-10-282-122A-64099	Sequence 64099, A
17	46	41.8	468 15	US-10-369-493-8063	Sequence 8063, Ap
18	45	40.9	205 12	US-10-456-479-8	Sequence 8, Appli
19	45	40.9	639 12	US-09-920-954-4	Sequence 4, Appli
20	44	40.0	583 9	US-09-758-269-2	Sequence 2, Appli
21	43.5	39.5	727 16	US-10-389-566-1335	Sequence 1335, Ap
22	43.5	39.5	727 16	US-10-389-566-2225	Sequence 2225, Ap
23	43	39.1	297 12	US-10-282-122A-45816	Sequence 45816, A
24	43	39.1	339 12	US-10-282-122A-48707	Sequence 48707, A
25	43	39.1	457 12	US-10-282-122A-55164	Sequence 55164, A
26	42	38.2	144 15	US-10-369-493-17305	Sequence 17305, A
27	42	38.2	484 12	US-10-282-122A-56323	Sequence 56323, A
28	42	38.2	963 12	US-10-424-599-194106	Sequence 194106, A
29	41	37.3	78 9	US-09-864-761-41675	Sequence 41675, A
30	41	37.3	270 9	US-09-733-569A-13	Sequence 13, Appl
31	41	37.3	270 15	US-10-427-218-13	Sequence 13, Appl
32	41	37.3	416 9	US-09-847-010-2	Sequence 2, Appli
33	41	37.3	416 14	US-10-235-939-2	Sequence 2, Appli
34	41	37.3	457 12	US-10-282-122A-67747	Sequence 67747, A
35	41	37.3	560 9	US-09-912-559-3	Sequence 3, Appli
36	41	37.3	560 9	US-09-912-559-4	Sequence 4, Appli
37	41	37.3	560 14	US-10-172-712-32	Sequence 32, Appl
38	41	37.3	560 15	US-10-391-215-5	Sequence 5, Appli
39	41	37.3	560 15	US-10-391-215-6	Sequence 6, Appli
40	41	37.3	560 15	US-10-391-215-7	Sequence 7, Appli
41	41	37.3	560 15	US-10-391-215-8	Sequence 8, Appli
42	41	37.3	688 15	US-10-369-493-6112	Sequence 6112, Ap
43	41	37.3	805 15	US-10-369-493-6498	Sequence 6498, Ap
44	41	37.3	805 15	US-10-369-493-6499	Sequence 6499, Ap
45	41	37.3	961 14	US-10-222-070-4	Sequence 4, Appli
46	41	37.3	1118 15	US-10-369-493-18845	Sequence 18845, A
47	40	36.4	48 9	US-09-847-010-21	Sequence 21, Appl
48	40	36.4	14 14	US-10-007-280A-214	Sequence 21, Appl
49	40	36.4	67 14	US-10-424-599-23554	Sequence 23554, A
50	40	36.4	82 12	US-09-764-881-5098	Sequence 5098, Ap
51	40	36.4	115 10	US-10-424-599-187049	Sequence 187049, Ap
52	40	36.4	179 12	US-10-282-122A-70025	Sequence 70025, A
53	40	36.4	210 12	US-10-424-599-194041	Sequence 194041, A
54	40	36.4	377 12	US-10-282-122A-69644	Sequence 69644, A
55	40	36.4	457 12	US-10-282-122A-48478	Sequence 48478, A
56	40	36.4	462 12	US-10-369-493-2027	Sequence 2027, Ap
57	40	36.4	594 15	US-10-389-566-2047	Sequence 2047, Ap
58	40	36.4	595 12	US-10-282-122A-51960	Sequence 51960, A
59	40	36.4	605 16	US-10-389-566-1405	Sequence 1405, Ap
60	40	36.4	888 16	US-10-389-566-1057	Sequence 1057, Ap
61	40	36.4	933 12	US-10-282-122A-47180	Sequence 47180, A
62	40	36.4	943 12	US-10-670-655-14	Sequence 14, Appl
63	40	36.4	1231 12	US-10-206-576-420	Sequence 420, App
64	40	36.4	1265 12	US-10-282-122A-57390	Sequence 57390, A
65	40	36.4	1265 12	US-10-206-576-418	Sequence 418, App
66	40	36.4	1648 9	US-09-515-806-4	Sequence 37, Appl
67	40	36.4	1648 10	US-09-847-758-37	Sequence 37, Appl
68	40	36.4	1648 12	US-10-174-333-37	Sequence 37, Appl
69	40	36.4	283 12	US-10-282-122A-69304	Sequence 69304, A
70	39.5	35.9	312 12	US-10-425-114-56959	Sequence 56959, A
71	39.5	35.9	312 12	US-10-425-114-56961	Sequence 56961, A
72	39.5	35.9	312 12	US-10-282-122A-73667	Sequence 73667, A
73	39.5	35.9	394 12	US-10-369-493-23374	Sequence 23374, A
74	39.5	35.9	445 15	US-10-424-599-178486	Sequence 178486, A
75	39.5	35.5	84 12	US-10-424-599-190450	Sequence 190450, A
76	39.5	35.5	89 12	US-10-424-599-144459	Sequence 144459, A
77	39.5	35.5	97 12	US-10-424-599-172541	Sequence 172541, A
78	39.5	35.5	119 12	US-10-424-599-264791	Sequence 264791, A
79	39.5	35.5	168 12	US-10-425-114-53155	Sequence 53155, A
80	39.5	35.5	182 12	US-10-282-122A-68016	Sequence 68016, A
81	39.5	35.5	191 12	US-09-978-244A-16	Sequence 16, Appli
82	39.5	35.5	234 10	US-10-425-114-69132	Sequence 69132, A
83	39.5	35.5	334 12	US-10-156-761-12953	Sequence 12953, A
84	39.5	35.5	357 14	US-09-738-626-3831	Sequence 3831, Ap
85	39.5	35.5	415 12	US-10-425-114-45668	Sequence 45668, A
86	39.5	35.5	431 12	US-10-425-114-53933	Sequence 53933, A
87	39.5	35.5			
88	39.5	35.5			

```
89 39 35.5 438 12 US-10-282-122A-46971 Sequence 46971, A
90 39 35.5 488 15 US-10-369-493-6826 Sequence 6826, Ap
91 39 35.5 509 9 US-09-258-031B-75 Sequence 75, App1
92 39 35.5 588 9 US-09-815-242-1253 Sequence 1253, A
93 39 35.5 588 9 US-09-815-242-1253 Sequence 1253, A
94 39 35.5 588 12 US-10-282-122A-44049 Sequence 44049, A
95 39 35.5 588 12 US-10-282-122A-70953 Sequence 70953, A
96 39 35.5 590 9 US-09-815-242-5344 Sequence 5344, Ap
97 39 35.5 612 12 US-10-012-600B-178 Sequence 178, App
98 39 35.5 642 15 US-10-369-493-13077 Sequence 13077, A
99 39 35.5 739 12 US-10-425-114-59903 Sequence 59903, A
100 39 35.5 763 12 US-10-424-599-184196 Sequence 184196, A
```

## ALIGNMENTS

```
RESULT 1
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. US20020155568A1
```

```
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIKHUIZEN, LOBBERT
; APPLICANT: RAHNOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10
```

```
Query Match 81.8%; Score 90; DB 9; Length 545;
Best Local Similarity 90.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 NYVYFIRAHDSVQTRIAKII 22
DB 156 NYIFIRAHDSVQTVIAKII 175
```

```
RESULT 2
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, SCOTT E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
```

```
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 81.8%; Score 90; DB 9; Length 1430;
Best Local Similarity 90.0%; Pred. No. 5.9e-06;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 NYVYFIRAHDSVQTRIAKII 22
DB 576 NYIFIRAHDSVQTVIAKII 595
```

```
RESULT 3
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. US20020155568A1
```

```
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIKHUIZEN, LOBBERT
; APPLICANT: RAHNOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11
```

```
Query Match 80.9%; Score 89; DB 9; Length 522;
Best Local Similarity 72.7%; Pred. No. 2.7e-06;
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 VNYVYFIRAHDSVQTRIAKII 22
DB 146 IPNYSFVRAHDSVQTVIAQIV 167
```

```
RESULT 4
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, SCOTT E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
```

;; PRIOR APPLICATION NUMBER: 09/008,172  
;; PRIOR FILING DATE: 1998-01-16  
;; PRIOR APPLICATION NUMBER: 08/482,711  
;; PRIOR FILING DATE: 1995-06-07  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 1475  
;; TYPE: PR1  
;; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 72.7%; Score 80; DB 9; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 0.00029;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPMYVIRAHDSVOTRIAKII 22  
DB 552 VPSYFIRAHDSVODLIRNII 573

RESULT 5  
US-09-740-274-4  
;; Sequence 4, Application US/09740274  
;; Patent No. US20020031826A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Nichols, Scott E.  
;; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
;; FILE REFERENCE: 0357CRD  
;; CURRENT APPLICATION NUMBER: US/09/740,274  
;; CURRENT FILING DATE: 2000-12-19  
;; PRIOR APPLICATION NUMBER: 09/210,361  
;; PRIOR FILING DATE: 1998-12-11  
;; PRIOR APPLICATION NUMBER: 09/007,999  
;; PRIOR FILING DATE: 1998-01-16  
;; PRIOR APPLICATION NUMBER: 08/478,704  
;; PRIOR FILING DATE: 1995-06-07  
;; PRIOR APPLICATION NUMBER: 09/009,620  
;; PRIOR FILING DATE: 1998-01-20  
;; PRIOR APPLICATION NUMBER: 08/485,243  
;; PRIOR FILING DATE: 1995-06-07  
;; PRIOR APPLICATION NUMBER: 09/008,172  
;; PRIOR FILING DATE: 1998-01-16  
;; PRIOR APPLICATION NUMBER: 08/482,711  
;; PRIOR FILING DATE: 1995-06-07  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 4  
;; LENGTH: 1375  
;; TYPE: PR1  
;; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 69.1%; Score 76; DB 9; Length 1375;  
Best Local Similarity 72.7%; Pred. No. 0.0013;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPMYVIRAHDSVOTRIAKII 22  
DB 578 VPSYFIRAHDSVODLIRNII 599

RESULT 6  
US-09-995-749A-13  
;; Sequence 13, Application US/09995749A  
;; Patent No. US20020155568A1  
;; GENERAL INFORMATION:  
;; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
;; APPLICANT: DIJKHUIZEN, LUBBERT  
;; APPLICANT: RAHAOUI, HAKIM  
;; APPLICANT: LEER, ROBERT-JAN  
;; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
;; FILE REFERENCE: B043388-CIP  
;; PRIOR APPLICATION NUMBER: EPO 00201871.1

;; CURRENT APPLICATION NUMBER: US/09/995,749A  
;; CURRENT FILING DATE: 2001-11-29  
;; PRIOR APPLICATION NUMBER: 09/604,957  
;; PRIOR FILING DATE: 2000-06-28  
;; PRIOR APPLICATION NUMBER: EPO 00201871.1  
;; PRIOR FILING DATE: 2000-05-25  
;; NUMBER OF SEQ ID NOS: 19  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO 13  
;; LENGTH: 535  
;; TYPE: PR1  
;; ORGANISM: Lactobacillus reuteri  
US-09-995-749A-13

Query Match 57.3%; Score 63; DB 9; Length 535;  
Best Local Similarity 50.0%; Pred. No. 0.064;  
Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPMYVIRAHDSVOTRIAKII 22  
DB 144 IPNYSFIRAHDDNNSODIOMAI 165

RESULT 7  
US-09-995-749A-12  
;; Sequence 12, Application US/09995749A  
;; Patent No. US20020155568A1  
;; GENERAL INFORMATION:  
;; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
;; APPLICANT: DIJKHUIZEN, LUBBERT  
;; APPLICANT: RAHAOUI, HAKIM  
;; APPLICANT: LEER, ROBERT-JAN  
;; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
;; FILE REFERENCE: B043388-CIP  
;; CURRENT APPLICATION NUMBER: US/09/995,749A  
;; CURRENT FILING DATE: 2001-11-29  
;; PRIOR APPLICATION NUMBER: 09/604,957  
;; PRIOR FILING DATE: 2000-06-28  
;; PRIOR APPLICATION NUMBER: EPO 00201871.1  
;; PRIOR FILING DATE: 2000-05-25  
;; NUMBER OF SEQ ID NOS: 19  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO 12  
;; LENGTH: 584  
;; TYPE: PR1  
;; ORGANISM: Leuconostoc mesenteroides  
US-09-995-749A-12

Query Match 57.3%; Score 63; DB 9; Length 584;  
Best Local Similarity 55.0%; Pred. No. 0.071;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPMYVIRAHDSVOTRIAK 20  
DB 167 IPNYSFIRAHDDYDAODPIRK 186

RESULT 8  
US-09-995-749A-2  
;; Sequence 2, Application US/09995749A  
;; Patent No. US20020155568A1  
;; GENERAL INFORMATION:  
;; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
;; APPLICANT: DIJKHUIZEN, LUBBERT  
;; APPLICANT: RAHAOUI, HAKIM  
;; APPLICANT: LEER, ROBERT-JAN  
;; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
;; FILE REFERENCE: B043388-CIP  
;; CURRENT APPLICATION NUMBER: US/09/995,749A  
;; CURRENT FILING DATE: 2001-11-29  
;; PRIOR APPLICATION NUMBER: 09/604,957  
;; PRIOR FILING DATE: 2000-06-28  
;; PRIOR APPLICATION NUMBER: EPO 00201871.1

PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1781  
TYPE: PRT  
ORGANISM: Lactobacillus reuteri  
US-09-995-749A-2

Query Match 57.3%; Score 63; DB 9; Length 1781;  
Best Local Similarity 50.0%; Pred. No. 0.26;  
Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 VNNYVFIKRAHDSVQTRIAKII 22  
DB 1123 IPNYSFVRADHDNNSQOQIQNAI 1144

RESULT 9  
US-10-417-280A-2  
Sequence 2, Application US/10417280A  
Publication No. US20030229923A1  
GENERAL INFORMATION:  
APPLICANT: KOSMANN, Jens  
APPLICANT: WELSH, Thomas  
APPLICANT: QUANZ, Martin  
APPLICANT: KNUTH, Karola  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
FILE REFERENCE: 0147-0247P  
CURRENT APPLICATION NUMBER: US/10/417,280A  
CURRENT FILING DATE: 2003-04-15  
PRIOR APPLICATION NUMBER: DE 19905069.4  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: US 09/499,203  
PRIOR FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-10-417-280A-2

Query Match 57.3%; Score 63; DB 15; Length 2057;  
Best Local Similarity 55.0%; Pred. No. 0.31;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VNNYVFIKRAHDSVQTRIAK 20  
DB 757 IPNYSFVRADHDYDAQDPPIRK 776

RESULT 10  
US-10-282-122A-76653  
Sequence 76653, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Foreyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 76653  
LENGTH: 602  
TYPE: PRT  
ORGANISM: Treponema pallidum  
US-10-282-122A-76653

Query Match 44.5%; Score 49; DB 12; Length 602;  
Best Local Similarity 56.2%; Pred. No. 17;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 IKAHDSVQTRIAKII 22  
DB 505 IRIHDYQLOKRIKIV 520

RESULT 11  
US-10-425-114-44862  
Sequence 44862, Application US/10425114  
Publication No. US2004003488A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(53113)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 44862  
LENGTH: 146  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700616705\_F11.pep  
US-10-425-114-44862

Query Match 43.6%; Score 48; DB 12; Length 146;  
Best Local Similarity 57.1%; Pred. No. 4.7;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 PNYSVFIKRAHDSVQ 15  
DB 1 PSYAVQKHSBVSQ 14

RESULT 12



```
US-10-425-114-61839
; Sequence 61839, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313) B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 7128
; SEQ ID NO 61839
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3227-021-C7_FLI.pep
US-10-425-114-61839

Query Match          43.6%; Score 48; DB 12; Length 170;
Best Local Similarity 57.1%; Pred. No. 5.6;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 PNYVFIKRAHDSVQ 15
DB 1 PSYVAQKHESEVQ 14

RESULT 13
US-10-424-599-279743
; Sequence 279743, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5323) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 279743
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(251)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_9462C.1.pep
US-10-424-599-279743

Query Match          43.6%; Score 48; DB 12; Length 251;
Best Local Similarity 64.3%; Pred. No. 8.8;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 YVFIRAHDSVQTR 17
DB 20 YLFAMAHDSVSTR 33

RESULT 14
US-10-369-493-2894
; Sequence 2894, Application US/10369493
; Publication No. US20030233675A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hunkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2894
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-369-493-2894

Query Match          43.6%; Score 48; DB 15; Length 492;
Best Local Similarity 36.4%; Pred. No. 19;
Matches 8; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNYVFIKRAHDSVQTRIAKII 22
DB 136 LPIKILMIRKHEPIYGIKISKIM 157

RESULT 15
US-10-282-122A-47139
; Sequence 47139, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

SEQ ID NO 47139  
 LENGTH: 506  
 TYPE: PRT  
 ORGANISM: Borrelia burgdorferi  
 US-10-282-122A-47139

Query Match 42.7%; Score 47; DB 12; Length 506;  
 Best Local Similarity 62.5%; Pred. No. 29;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 IRAHSEVQTRIAKI 22  
 DB 406 IRIHNELEQRIFKII 421

RESULT 16  
 US-10-282-122A-64099  
 ; Sequence 64099, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Foreyth, R.  
 APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

RESULT 17  
 US-10-369-493-8063

; Sequence 8063, Application US/10369493  
 ; Publication No. US2003023675A1  
 ; GENERAL INFORMATION:

APPLICANT: Cao, Yongwei  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Chen, Xiandeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT FILING DATE: 2003-02-28

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 8063

LENGTH: 468

TYPE: PRT

ORGANISM: Rhodobacter sphaeroides

US-10-369-493-8063

Query Match 41.8%; Score 46; DB 15; Length 468;  
 Best Local Similarity 28.6%; Pred. No. 39;  
 Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 PNYYFIRAHDSVQTRIAKI 22  
 DB 128 PRLVWVEHREPAIRERNAKVL 148

RESULT 18  
 US-10-456-479-8

; Sequence 8, Application US/10456479  
 ; Publication No. US20040072321A1  
 ; GENERAL INFORMATION:

APPLICANT: SATO, TSUYOSHI

APPLICANT: OKUDA, MITSUYOSHI

APPLICANT: TAKIMURA, YASUOSHI

APPLICANT: SUMITOMO, NOBUYUKI

APPLICANT: NOMURA, MASATUMI

APPLICANT: KOBAYASHI, TORU

TITLE OF INVENTION: ALKALINE PROTEASE

FILE REFERENCE: 238700USO

CURRENT FILING DATE: 2003-06-09

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

PRIOR FILING DATE: 2002-06-26

1 Publication No. US20020064854A1  
 2 GENERAL INFORMATION:  
 3 APPLICANT: TAKATAMA, MIKIO  
 4 APPLICANT: OKUDA, MITSUYOSHI  
 5 APPLICANT: SAKKI, KATSUYOSHI  
 6 APPLICANT: KUBOTA, HIROMI  
 7 APPLICANT: HITOMI, JUN  
 8 APPLICANT: KAGEYAMA, YASUSHI  
 9 APPLICANT: SHIKATA, SHITSUW  
 10 APPLICANT: NOMURA, MASAFUMI  
 11 TITLE OF INVENTION: ALKALINE PROTEASE  
 12 FILE REFERENCE: 0327-0832-0PCT  
 13 CURRENT APPLICATION NUMBER: US/09/920,955  
 14 CURRENT FILING DATE: 2001-08-03  
 15 PRIOR APPLICATION NUMBER: 09/509,814  
 16 PRIOR FILING DATE: 2000-04-06  
 17 PRIOR APPLICATION NUMBER: PCT/JP98/04528  
 18 PRIOR FILING DATE: 1998-10-07  
 19 PRIOR APPLICATION NUMBER: JP 9-274570  
 20 PRIOR FILING DATE: 1997-06-08  
 21 NUMBER OF SEQ ID NOS: 24  
 22 SOFTWARE: PatentIn version 3.0  
 23 SEQ ID NO 4  
 24 LENGTH: 639  
 25 TYPE: PRT  
 26 ORGANISM: Bacillus sp.  
 27 US-09-920-954-4

Query Match	40.9%;	Score 45;	DB 12;	Length 639;
Best Local Similarity	28.6%;	Pred. NO. 83;		
Matches	6;	Conservative	9;	Mismatches 6;
				Indels 0;
				Gaps 0;

Qy 1 VPNYVFIRAHMSEVQTRI AKI 21  
 :|:|:|:|:|:|:|  
 Db 114 IPDYAYIVEYEGDVQSKVR SI 134

```

RESULT 20
US-09-758-269-2
; Sequence 2, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLARVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 583
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
; US-09-758-269-2

```

Query Match	40.0%;	Score 44;	DB 9;	Length 583;
Best Local Similarity	42.1%;	Pred. No. 1.1e+02;		
Matches	8;	Conservative	3;	Mismatches 8;
				Indels 0;
				Gaps 0;

```
QY      4 YVFIRAHDSVQTRIAKII 22
        |::|||::||
Db      530 YIFCHVHDEETKTSSELQII 548
```

RESULT 21  
US-10-389-566-1335

```

: Sequence 1335, Application US/10389566
: Publication No. US20040025202A1
: GENERAL INFORMATION:
: APPLICANT: Monsanto Technology, LLC
: APPLICANT: Laurie, Cathy C
: TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
: FILE REFERENCE: 38-77(52900)D
: CURRENT APPLICATION NUMBER: US/10/389,566
: CURRENT FILING DATE: 2003-03-31
: PRIOR APPLICATION NUMBER: US 60/365,301
: PRIOR FILING DATE: 2002-03-15
: PRIOR APPLICATION NUMBER: US 60/391,786
: PRIOR FILING DATE: 2002-06-25
: PRIOR APPLICATION NUMBER: US 60/392,018
: PRIOR FILING DATE: 2002-06-26
: NUMBER OF SEQ ID NOS: 2459
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 1335
: LENGTH: 727
: TYPE: PRT
: ORGANISM: Physcomitrella patens
: US-10-389-566-1335

```

Query Match	39.5%	Score 43.5;	DB 16;	Length 727;
Best Local Similarity	45.5%;	Pred. No. 1.7e+02;		
Matches 10; Conservative	4;	Mismatches 7;	Indels 1;	Gaps 1

QY 1 VPNTVFIRAHDSVQTRIAKI 22  
Db 553 VPNQFIRAHDS-IMNRSAMV 5733

```

RESULT 22
US-10-389-566-2225
? Sequence 2225, Application US/10389566
? Publication No. US20040025202A1
? GENERAL INFORMATION:
? APPLICANT: Monsanto Technology, LLC
? APPLICANT: Laurie, Cathy C
? TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
? FILE REFERENCE: 38-77(52900)D
? CURRENT APPLICATION NUMBER: US/10/389,566
? CURRENT FILING DATE: 2003-03-31
? PRIOR APPLICATION NUMBER: US 60/365,301
? PRIOR FILING DATE: 2002-03-15
? PRIOR APPLICATION NUMBER: US 60/391,786
? PRIOR FILING DATE: 2002-06-25
? PRIOR APPLICATION NUMBER: US 60/392,018
? PRIOR FILING DATE: 2002-06-26
? NUMBER OF SEQ ID NOS: 2459
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 2225
? LENGTH: 727
? TYPE: PRT
? ORGANISM: Physcomitrella patens
? US-10-389-566-2225

```

Query Match	39.5%	Score 43.5;	DB 16;	length 727;
Best Local Similarity	45.5%;	Pred. No. 1.7e+02;		
Matches 10;	Conservative 4;	Mismatches 7;	Indels 1;	Gaps 1;

QY 1 VPNTVFIRAHDSVQTRIAKII 22  
||| ||| : | : :  
Db 553 VPNQFIRAHDS-IMNRSAAVV 573

RESULT 23  
US-10-282-122A-45616  
; Sequence 45616, Application US/10282122A1  
; Publication No. US2004002122A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos

```

; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45816
; LENGTH: 297
; TYPE: PRF
; ORGANISM: Bacillus anthracis
; US-10-282-122A-45816

Query Match          39.1%; Score 43; DB 12; Length 297;
Best Local Similarity 41.2%; Pred. No. 74;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY      6 FIRAHDSVQTRIAKI 22
DB      272 FVRVHDVKEMSRMAKMM 288

RESULT 24
; US-10-282-122A-48707
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/253,625
```

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48707
; LENGTH: 339
; TYPE: PRF
; ORGANISM: Bacteroides fragilis
; US-10-282-122A-48707

Query Match          39.1%; Score 43; DB 12; Length 339;
Best Local Similarity 41.2%; Pred. No. 86;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      5 VFIRAHDSVQTRIAKI 21
DB      161 VILRHTSSVQSRVMEV 177

RESULT 25
; US-10-282-122A-55164
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 55164  
LENGTH: 457  
TYPE: PRT  
ORGANISM: Chlamydia trachomatis  
US-10-282-122A-55164

Query Match 39.1%; Score 43; DB 12; Length 457;  
Best Local Similarity 45.0%; Pred. No. 1.2e+02;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 NNYFIRAHDSVQTRIAKII 22  
DB 219 NNYVIRGERDCTVQRRQKLI 238

RESULT 26  
US-10-369-493-17305

Sequence 17305, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 17305  
LENGTH: 144  
TYPE: PRT  
ORGANISM: Bacillus halodurans  
US-10-369-493-17305

Query Match 38.2%; Score 42; DB 15; Length 144;  
Best Local Similarity 61.5%; Pred. No. 47;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PNYFIRAHDSV 14  
DB 74 PNYVIRIRIHEEV 86

RESULT 27

US-10-282-122A-56323  
Sequence 56323, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.

APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EUTRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 56323  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Enterobacter cloacae  
US-10-282-122A-56323

Query Match 38.2%; Score 42; DB 12; Length 484;  
Best Local Similarity 23.8%; Pred. No. 1.9e+02;  
Matches 5; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 2 PNYFIRAHDSVQTRIAKII 22  
DB 130 PKLWVGRHEPRFQVAKVL 150

RESULT 28

US-10-424-599-194106  
Sequence 194106, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 194106  
LENGTH: 963  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(963)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_17303C.1.pep  
US-10-424-599-194106

Query Match 38.2%; Score 42; DB 12; Length 963;  
Best Local Similarity 46.7%; Pred. No. 4.3e+02;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 PNVFIRAHDSVQRT 16  
Db 774 PXYLIAMVHDSVST 788

RESULT 29  
US-09-864-761-41675  
Sequence 41675, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 41675  
LENGTH: 78  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC007221.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2  
OTHER INFORMATION: EST\_HUMAN HIT: BE788578.1, EVALUATE 2.90e+00

US-09-864-761-41675

Query Match 37.3%; Score 41; DB 9; Length 78;  
Best Local Similarity 38.9%; Pred. No. 34;  
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 4 YVFIARHDSVQRTAKI 21  
Db 38 YVFIARHDSVQRTAKI 55

RESULT 30  
US-09-733-569A-13  
Sequence 13, Application US/09733569A  
Patent No. US20020045739A1  
GENERAL INFORMATION:  
APPLICANT: Regents of the University of California  
APPLICANT: Gerald, NEWTON  
APPLICANT: Av-Gay, YOSSEF  
APPLICANT: Robert, FAHY  
TITLE OF INVENTION: ACT1 GLUCOSAMINYL INOSITOL AMIDASE FAMILY AND METHODS OF USE  
FILE REFERENCE: UCS01170-1  
CURRENT APPLICATION NUMBER: US/09/733,569A  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,503  
PRIOR FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 13  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Streptomyces lincolnensis  
US-09-733-569A-13

Query Match 37.3%; Score 41; DB 9; Length 270;  
Best Local Similarity 47.1%; Pred. No. 1.4e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 FVRAPIRABATRLARVI 22  
Db 100 FVRAPIRABATRLARVI 116

Search completed: May 4, 2004, 09:25:42  
Job time : 37.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:02:05 ; Search time 47.6667 Seconds  
(without alignments)  
130.407 Million cell updates/sec

Title: US-09-290-049a-19

Perfect score: 110  
Sequence: 1 VPSYSPARAHSEVDIIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 100 summaries

Database: A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	110	100.0	1590	7	ADD93657 Streptoc
2	110	100.0	1592	2	AA832925
3	103	93.6	1017	5	AAU79285
4	103	93.6	1476	5	AAU79284
5	98	89.1	1375	5	AAU98028
6	98	89.1	1375	5	AAU79288
7	98	89.1	1375	5	ADD93655
8	97	88.2	1475	5	AAU98031
9	97	88.2	1475	5	AAU98040
10	97	88.2	1475	5	AAU98033
11	97	88.2	1475	5	AAU98030
12	97	88.2	1475	5	AAU98039
13	97	88.2	1475	5	AAU98027
14	97	88.2	1475	5	ADD93654
15	90	81.8	1475	5	AAU98035
16	90	81.8	1475	5	AAU98034
17	90	81.8	1475	5	AAU98032
18	86	78.2	2055	6	ABR63235
19	83	75.5	1149	6	ABR63236
20	83	75.5	1475	5	AAU98036
21	83	75.5	1475	5	AAU98037
22	83	75.5	1475	5	AAU98038
23	82	74.5	1223	6	ABR63228
24	81	73.6	1499	7	ADC54806
25	81	73.6	1527	5	AAU80055

26	81	73.6	1527	7	ADC54807	ADC54807 Leuconost
27	80	72.7	1554	7	ADD93656	ADD93656 Streptoc
28	79	71.8	221	6	ABR63229	ABR63229 Glucanuc
29	76	69.1	224	6	ABR63227	ABR63227 Glucanuc
30	76	69.1	1497	6	ABR63234	ABR63234 Glucanuc
31	75	68.2	221	6	ABR63226	ABR63226 Glucanuc
32	75	68.2	223	6	ABR63230	ABR63230 Glucanuc
33	75	68.2	2147	6	ABR63231	ABR63231 Glucanuc
34	75	68.2	2835	5	ABR88574	ABR88574 Dextran s
35	75	68.2	2835	5	ABR88574	ABR88574 Dextran s
36	74	67.3	15	5	ABR98651	ABR98651 Dextrane-
37	71	64.5	1430	5	AAU98044	AAU98044 S. mutans
38	71	64.5	1430	5	AAU98041	AAU98041 S. mutans
39	69	62.7	15	5	ABR98650	ABR98650 Dextrane-
40	69	62.7	1365	7	ADD93659	ADD93659 Streptoc
41	68	61.8	1781	5	AAU74519	AAU74519 Lactobaci
42	68	61.8	2057	3	AAU10667	AAU10667 L. mesent
43	67	60.9	1430	5	AAU98045	AAU98045 S. mutans
44	67	60.9	1430	5	AAU98042	AAU98042 S. mutans
45	67	60.9	1518	7	ADD93660	ADD93660 Streptoc
46	64	58.2	15	5	ABR98654	ABR98654 Dextrane-
47	64	58.2	15	5	ABR98655	ABR98655 Dextrane-
48	64	58.2	15	5	ABR98653	ABR98653 Dextrane-
49	64	58.2	1430	5	AAU98043	AAU98043 S. mutans
50	64	58.2	1430	5	AAU98029	AAU98029 S. mutans
51	64	58.2	1430	5	ADD93656	ADD93656 Streptoc
52	64	58.2	1577	2	AAU91047	AAU91047 Alpha-D-g
53	62	56.4	2022	6	ABR63232	ABR63232 Glucanuc
54	62	54.5	15	5	ABR98657	ABR98657 Dextrane-
55	60	54.5	15	5	ABR98579	ABR98579 Dextrane-
56	60	54.5	15	5	ABR85559	ABR85559 Peptide d
57	52	47.3	15	5	ABR98656	ABR98656 Dextrane-
58	51	46.4	15	5	ABR98652	ABR98652 Dextrane-
59	48.5	44.1	465	6	ABU43776	ABU43776 Protein e
60	47.5	43.2	481	4	AAU34088	AAU34088 Staphyloc
61	47.5	43.2	487	4	AAU36863	AAU36863 Staphyloc
62	47	42.7	214	4	ABG01221	ABG01221 Novel hum
63	47	42.7	881	5	ABP65773	ABP65773 Bifidobac
64	46.5	42.3	484	6	ABU43150	ABU43150 Protein e
65	46.5	42.3	501	6	ABP39270	ABP39270 Staphyloc
66	46	41.8	479	4	ABG20960	ABG20960 Novel hum
67	46	41.8	490	4	ABU11733	ABU11733 Human bre
68	46	41.8	519	2	AAU48586	AAU48586 Human bre
69	46	41.8	724	4	AAU79198	AAU79198 Human pro
70	46	41.8	743	4	AAU93069	AAU93069 Human pro
71	46	41.8	743	4	AAU93288	AAU93288 Human pro
72	46	41.8	743	5	AAU15259	AAU15259 Human RNA
73	46	41.8	779	5	ABR97147	ABR97147 Human tum
74	46	41.8	1345	5	AAU825097	AAU825097 Human kin
75	46	41.8	2193	6	ABR42219	ABR42219 Human kin
76	45.5	41.4	484	2	AAU21897	AAU21897 Glutamy-
77	45.5	41.4	484	6	ABU15879	ABU15879 Protein e
78	45.5	41.4	487	6	ABU72476	ABU72476 Staphyloc
79	45	40.9	141	4	AAU09497	AAU09497 Human pol
80	45	40.9	141	7	ADC33073	ADC33073 Human nov
81	45	40.9	255	5	AAU980387	AAU980387 Human lun
82	45	40.9	255	6	ABP98904	ABP98904 Human mol
83	45	40.9	255	7	ADC31420	ADC31420 Human nov
84	45	40.9	255	7	ADB95126	ADB95126 Human lun
85	45	40.9	262	5	AAU80388	AAU80388 Human lun
86	45	40.9	262	7	ADB95129	ADB95129 Human lun
87	45	40.9	541	1	AAU70433	AAU70433 Human atr
88	45	40.9	541	2	AAU95586	AAU95586 Atrial na
89	45	40.9	597	6	ABU48783	ABU48783 Protein e
90	45	40.9	704	6	ABR80170	ABR80170 A. funiga
91	45	40.9	855	5	ABR89573	ABR89573 Dextran s
92	44.5	40.5	386	7	ADB82758	ADB82758 Human pro
93	44.5	40.5	451	4	AAU41255	AAU41255 Human pol
94	44.5	40.5	486	4	AAU37862	AAU37862 Streptoc
95	44.5	40.5	486	6	ABU46275	ABU46275 Protein e
96	44.5	40.5	603	5	ABR48333	ABR48333 Listeria
97	44	40.0	216	5	ABR48421	ABR48421 Listeria
98	44	40.0	527	5	ABR80879	ABR80879 Maize cyt

99 44 40.0 565 7 ADC94590  
100 44 40.0 722 4 ABG28353

Adc94590 E. faeciu  
Abg28353 Novel hum

## ALIGNMENTS

RESULT 1  
ADD93657

ID ADD93657 standard; protein; 1590 AA.

AC ADD93657;

DT 29-JAN-2004 (first entry)

DE Streptococcus sobrinus glucosyltransferase-I.

KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

OS Streptococcus sobrinus.

PN MO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006962.

PR 07-MAR-2003; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PA (FORS-) FORSYTH INST.

PI Smith DJ, Taubman MA;

DR WPI; 2003-845091/78.

PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

PS Claim 16; Page 14; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus sobrinus  
XX glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially  
XX from the catalytic domain of the polypeptide, can be used in immunogenic  
XX compositions and subunit vaccines for dental caries. These compositions  
XX comprise a major histocompatibility complex (MHC) class II protein-  
XX binding peptide from S. mutans glucan binding protein-B (GbpB)  
XX covalently linked with a peptide fragment of a streptococcal  
XX glucosyltransferase. The compositions are used in a claimed method of  
XX eliciting production of an antibody in a mammal. Depliptic or  
XX multipletopic polypeptides can be prepared synthetically or by  
XX recombinant DNA technology. Antibodies raised against MHC class II  
XX binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1590 AA;

Query Match 100.0%; Score 110; DB 7; Length 1590;

Best Local Similarity 100.0%; Pred. No. 1.9e-09; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVDIIRDI 22

DB 548 VPSYSPARAHSEVDIIRDI 569

RESULT 2

ID AAR32925 standard; protein; 1592 AA.

XX AAR32925;

DT 28-JUN-1993 (first entry)

XX Glucosyltransferase I.

DE GT-1; Streptococcus; dental; caries.

OS Streptococcus sobrinus.

PN JP05023188-A.

PD 02-FEB-1993.

PF 25-JUL-1991; 91JP-00186592.

PR 25-JUL-1991; 91JP-00186592.

PA (KATO/) KATO K.

PA (FUKU/) FUKU I.

DR WPI; 1993-079449/10.

DR N-PSDB; AAQ37760.

PT DNA sequence glucosyl:transferase-I - comprises Streptococcus sobrinus

PS Claim 13; Page 15; 29pp; Japanese.

XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
XX glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.  
XX sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially  
XX digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp  
XX fragment was ligated into pUC18 and E. coli JM109 transformed with it. A  
XX GT-1 expressing clone was isolated and sequenced. The clone may be used  
XX in the development of a drug for dental caries

SQ Sequence 1592 AA;

Query Match 100.0%; Score 110; DB 2; Length 1592;

Best Local Similarity 100.0%; Pred. No. 1.9e-09; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVDIIRDI 22

DB 548 VPSYSPARAHSEVDIIRDI 569

RESULT 3

ID AAU79285 standard; protein; 1017 AA.

XX AAU79285;

DT 13-AUG-2002 (first entry)

DE Streptococcus mutans monoclonal antibody-related protein #2.

XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;

XX glucosyl transferase-B; immunotherapy.

OS Streptococcus mutans.

PN JP2002114709-A.

PD 16-APR-2002.

PF 04-OCT-2000; 2000JP-00304889.

PR 04-OCT-2000; 2000JP-00304889.

PA (UNIV-) UNIV NIPPON.

DR WPI; 2002-448885/48.

PT Anti-carries agent composed of a monoclonal antibody against an inhibitory



PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 XX  
 PS Claim 4; Page 17-19; 28pp; Japanese.  
 CC The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein  
 CC  
 SQ Sequence 1017 AA;  
 Query Match 93.6%; Score 103; DB 5; Length 1017;  
 Best Local Similarity 90.9%; Pred. No. 1.7e-08;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VPSYSPARHDSVODLIRDI 22  
 DB 518 VPSYSPARHDSVODLIRDI 539  
 RESULT 4  
 AAU79284  
 ID AAU79284 standard; protein; 1476 AA.  
 AC AAU79284;  
 XX  
 XX 13-AUG-2002 (first entry)  
 DT Streptococcus mutans monoclonal antibody-related protein #1.  
 DE Streptococcus mutans monoclonal antibody-related protein #1.  
 XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
 KM glucosyl transferase-B; immunotherapy.  
 XX Streptococcus mutans.  
 OS Streptococcus mutans.  
 XX  
 PN JP2002114709-A.  
 PD 16-APR-2002.  
 XX  
 PF 04-OCT-2000; 2000JP-00304889.  
 XX  
 PR 04-OCT-2000; 2000JP-00304889.  
 XX  
 PA (UYNI-) UNIV NIPPON.  
 XX  
 XX WPI; 2002-448885/48.  
 DR  
 XX  
 PT Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 XX  
 PS Claim 3; Page 13-16; 28pp; Japanese.  
 CC The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein  
 CC  
 SQ Sequence 1476 AA;

Query Match 93.6%; Score 103; DB 5; Length 1476;  
 Best Local Similarity 90.9%; Pred. No. 2.7e-08;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VPSYSPARHDSVODLIRDI 22  
 DB 552 VPSYSPARHDSVODLIRDI 573  
 RESULT 5  
 AAU98028  
 ID AAU98028 standard; protein; 1375 AA.  
 AC AAU98028;  
 XX  
 XX 27-AUG-2002 (first entry)  
 DT S. mutans glucosyltransferase GTFC.  
 DE  
 XX  
 XX Glucosyltransferase; GTFC; transgenic plant; paper sizing;  
 KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture.  
 XX  
 XX Streptococcus mutans.  
 OS  
 XX  
 XX US2002031826-A1.  
 PN  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX  
 PI Nichols SE;  
 XX  
 XX WPI; 2002-414332/44.  
 DR N-FSDB; ABR52939.  
 DR  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Disclosure; Page 30-33; 44pp; English.  
 XX  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from 1589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilizing the glucan produced by GTF, which utilizes  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents GTFc  
XX

SQ Sequence 1375 AA;

Query Match 89.1%; Score 98; DB 5; Length 1375;  
Best Local Similarity 86.4%; Pred. No. 1.8e-07;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDIIRNII 22  
|||||  
578 VPSYSPIRAHDSVQDLIRNII 599

RESULT 6  
AAU79288

ID AAU79288 standard; protein; 1375 AA.

XX AAU79288;

XX 13-AUG-2002 (first entry)

DE Streptococcus mutans monoclonal antibody-related protein #5.

XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
KM glucosyl transferase-B; immunotherapy.

XX Streptococcus mutans.

XX JP2002114709-A.

XX 16-APR-2002.

XX 04-OCT-2000; 2000JP-00304889.

XX 04-OCT-2000; 2000JP-00304889.

XX (UYNI-) UNIV NIPPON.

XX WPI; 2002-448885/48.

XX Anti-carries agent composed of a monoclonal antibody against an inhibitory  
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
PT B (GTF-B) of Streptococcus mutans.

XX Disclosure; Page 22-25; 28pp; Japanese.

XX The invention relates to a monoclonal antibody against dental caries and  
CC an anti-carries agent composed of a monoclonal antibody produced by  
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
CC transferase-B. The monoclonal antibody specifically inhibits water  
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
CC transferase-B and is used in the immunotherapy of dental caries. This  
CC sequence represents a Streptococcus mutans monoclonal antibody-related  
CC protein  
XX

SQ Sequence 1375 AA;

Query Match 89.1%; Score 98; DB 5; Length 1375;

Best Local Similarity 86.4%; Pred. No. 1.8e-07;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDIIRNII 22  
|||||  
578 VPSYSPIRAHDSVQDLIRNII 599

RESULT 7  
ADD93655

ID ADD93655 standard; protein; 1375 AA.

XX ADD93655;

XX 29-JAN-2004 (first entry)

DE Streptococcus mutans glucosyltransferase-C.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX

XX Claim 16; Page 13; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus mutans  
CC glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially  
CC from the catalytic domain of the polypeptide, can be used in immunogenic  
CC compositions and subunit vaccines for dental caries. These compositions  
CC comprise a major histocompatibility complex (MHC) class II protein-  
CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
CC covalently linked with a peptide fragment of a streptococcal  
CC glucosyltransferase. The compositions are used in a claimed method of  
CC eliciting production of an antibody in a mammal. Dieptopic or  
CC multipitopic polypeptides can be prepared synthetically or by  
CC recombinant DNA technology. Antibodies raised against MHC class II  
CC binding fragments of GbpB can be used in passive immunisation.  
XX

SQ Sequence 1375 AA;

Query Match 89.1%; Score 98; DB 7; Length 1375;  
Best Local Similarity 86.4%; Pred. No. 1.8e-07;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDIIRNII 22  
|||||  
578 VPSYSPIRAHDSVQDLIRNII 599

RESULT 8  
AAU98031

ID AAU98031 standard; protein; 1475 AA.

XX AAU98031;

XX

DT 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB mutant D457N.  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FT Misc-difference 457  
 FT /note= "Wild-type Asp substituted by Asn"  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 PI MPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture, unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX Claim 36  
 XX Sequence 1475 AA;  
 SQ  
 Query Match 88.2%; Score 97; DB 5; Length 1475;  
 Best Local Similarity 86.4%; Pred. No. 2.9e-07;  
 Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 VPSSFPARADSEYODIIRDI 22  
 Db 552 VPSSFPARADSEYODIIRDI 573  
 RESULT 9  
 AAU98040  
 ID AAU98040 standard; protein; 1475 AA.  
 XX AAU98040;  
 AC  
 XX  
 AC AAU98040;  
 XX  
 AC  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB mutant K779Q.  
 DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 KM amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FT Misc-difference 779  
 FT /note= "Wild-type Lys substituted by Gln"  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 PI MPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX SQ Sequence 1475 AA;  
 CC  
 CC Query Match 88.2%; Score 97; DB 5; Length 1475;  
 CC Best Local Similarity 86.4%; Pred. No. 2.9e-07;  
 CC Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSYSFARADSEVODIIRDI 22  
 DB 552 VPSYSFIRADSEVODIIRDI 573

RESULT 10  
 AAU98033  
 ID AAU98033 standard; protein; 1475 AA.  
 XX AC  
 XX AAU98033;  
 DT 27-AUG-2002 (first entry)

XX DE 5. mutants glucosyltransferase GTFB mutant KI014T.  
 XX KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX KM amyloplast; vacuole; paper manufacture; mutant; muteln.  
 XX OS Streptococcus mutans.  
 XX OS Synthetic.  
 XX FT Key Location/Qualifiers  
 FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"  
 FT US2002031826-A1.  
 PD 14-MAR-2002.  
 XX PF 19-DEC-2000; 2000US-00740274.  
 XX PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX PI Nichols SE;  
 XX DR WPI; 2002-414332/44.  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC KI014T, D457N/D567T, D457N/D571K, D567T/D571K/KI014T,  
 CC I448V/D457N/D567T/D571K/K779Q/KI014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX SQ Sequence 1475 AA;  
 CC  
 CC Query Match 88.2%; Score 97; DB 5; Length 1475;  
 CC Best Local Similarity 86.4%; Pred. No. 2.9e-07;  
 CC Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSYSFARADSEVODIIRDI 22  
 DB 552 VPSYSFIRADSEVODIIRDI 573

RESULT 11  
 AAU98030  
 ID AAU98030 standard; protein; 1475 AA.  
 XX AC  
 XX AAU98030;  
 DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant I448V.  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutcin.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FT Misc-difference 448 /note= "Wild-type Ile substituted by Val"  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 XX  
 XX Query Match 88.2%; Score 97; DB 5; Length 1475;  
 XX Best Local Similarity 86.4%; Pred. No. 2.9e-07;  
 XX Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 XX QY 1 VPSYSPARADSEYVDIIRDI 22  
 XX |||||  
 XX Db 552 VPSYSPARADSEYVDIIRDI 573  
 XX  
 XX RESULT 12  
 XX AAU98039  
 XX ID AAU98039 standard; protein; 1475 AA.  
 XX AC AAU98039;  
 XX XX  
 XX 27-AUG-2002 (first entry)  
 XX DE S. mutans glucosyltransferase GTFB mutant YYY169-171AAA.  
 XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutcin.  
 XX OS Streptococcus mutans.  
 XX Synthetic.  
 XX Key Location/Qualifiers  
 FT Misc-difference 169; .171  
 FT /note= "Wild-type Tyr-Tyr-Tyr substituted by Ala-Ala-Ala"  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked a promoter, a vector comprising the expression cassette, host  
CC cell introduced with the vector, a transgenic plant comprising the  
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
CC coating composition comprising a glucan produced in a plant transformed  
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
CC thermoplastic molecule or their combinations or glucan and starch where  
CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilising the glucan produced by GTF, which utilises  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTFB mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the index  
CC using the GTFB sequence appearing as AA098027 and the information in  
CC claim 36

XX Sequence 1475 AA;

50

```

Query Match Similarity      88.2%; Score 97; DB 5; Length 1475;
Best Local Similarity      86.4%; Pred. No. 2.9e-07;
Matches      19; Conservative      1; Mismatches      2; Indels      0; Gaps      0

OY      1 VPSYSFARAHSEVDIIIRDTI 22
      |||||:|||||:|||||
Db      552 VPSYSFIRAHSEVDLIADII 573

```

RESULT 13  
AAU98027  
ID AAU98027 standard; protein; 1475 AA  
XX  
AC AAU98027;

DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB.  
 XX  
 KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KM coating composition; glucan; starch; latex; thermoplastic molecule  
 KW amyloplast; vacuole; paper manufacture.  
 XX  
 OS Streptococcus mutans.

PN US2002031826-A1.

PD 14-MAR-2002

PF 19-DEC-2000; 2000US-00740274.

PR 07-JUN-1995; 95US-00478704.  
DT TNY 1995; 95US-00483711  
SETTS-00483711

PR 07-JUN-1995; 95US-00485243;  
15 JAN 1999; 98US-00007999

PR 16-JAN-1998; 98US-000081/2  
 DP 30-JAN-1998; 98US-00009620

PR 11-DEC-1998; 9805-00210361  
XX

PA (NICH/) NICHOLS S E

XX

PI Nichols SE;

DR WPI; 2002-414332/44.

XX

PT Glucosyltransferase B or D protein useful for producing a glucan useful  
PT as substitutes for and additions to modified starch and latexes in paper  
PT manufacture, comprises mutations in specific positions.

PS Disclosure; Page 21-25; 44pp; English

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448T, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K/X1014T, D457N/D571K/D567T/D571K/K719Q/K1014T, Y169A/Y170A/Y171A, and K719Q or a GTF B polypeptide having changes at positions from T589D, T589E, N471D, M471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide, a ribonucleic acid sequence encoding the complementary polynucleotide, a transgenic plant comprising the operably linked GTF mutant, an expression cassette comprising the expression cassette, host cell linked to a promoter, a vector comprising the expression cassette, a cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermo-plastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a male line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising a glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermo-plastic properties and impart gloss to the paper during coating step. The present sequence represents GTFB

**SQ Sequence 1475 AA:**

Query Match	88.2%;	Score 97;	DB 5;	Length 1475;
Best Local Similarity	86.4%;	Pred. NO. 2.9e-07;		
Matches 15; Conservative		1; Mismatches 2;	Indels 0;	Gaps 0

QY 1 VPSTSFARAHADSEVQDIIRDII 22  
||| ||| ||| ||| : ||| |||  
Db 552 VPSYSFIRAHADSEVQDLIADI 573

## RESULT 14

ID ADD93654 standard; protein; 1475 AA

AC ADD93654

DT 29-JAN-2004 (first entry)

DE Streptococcus mutans glucosyltransferase-B.

KW Glucosyltransferase; enzyme; vaccine; anticaries; epilope; immunogen...

OS Streptococcus mutans.

XX  
PN  
W02003075845-A2

FA

XX 18-SEP-2003.  
 XX 07-MAR-2003; 2003MO-US006962.  
 XX 07-MAR-2002; 2002US-0363209P.  
 XX 08-AUG-2002; 2002US-0402483P.  
 XX (FORS-) FORSYTH INST.  
 XX Smith DJ, Taudman MA;  
 XX WPI; 2003-845091/78.  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX Claim 16; Page 12-13; 49pp; English.  
 XX The present sequence is the protein sequence of Streptococcus mutans  
 CC glucosyltransferase-B (GTF-B). Peptide fragments of GTF-B, especially  
 CC from the catalytic domain of the polypeptide, can be used in immunogenic  
 CC compositions and subunit vaccines for dental caries. These compositions  
 CC comprise a major histocompatibility complex (MHC) class II protein-  
 CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
 CC covalently linked with a peptide fragment of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Depliptic or  
 CC multi-epitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX Sequence 1475 AA;  
 SQ  
 Query Match 88.2%; Score 97; DB 7; Length 1475;  
 Best Local Similarity 86.4%; Pred. No. 2.9e-07;  
 Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 VPSYSPARADSEVODIIRDI 22  
 DB 552 VPSYSPIRADSEVODIIRDI 573  
 RESULT 15  
 AAU98035  
 ID AAU98035 standard; protein; 1475 AA.  
 XX AAU98035;  
 XX 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX Streptococcus mutans.  
 XX Synthetic.  
 XX Key Location/Qualifiers  
 XX Misc-difference 457  
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Asn"  
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX

PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, or a  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q,  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the expression cassette, host  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole of a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing the glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll a/b binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 SQ Sequence 1475 AA;  
 Query Match 81.8%; Score 90; DB 5; Length 1475;  
 Best Local Similarity 81.8%; Pred. No. 4.5e-06;  
 Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 VPSYSPARADSEVODIIRDI 22  
 DB 552 VPSYSPIRADSEVODIIRDI 573  
 RESULT 16  
 AAU98034  
 ID AAU98034 standard; protein; 1475 AA.  
 XX



XX AC AAU98034;  
 XX DT 27-AUG-2002 (first entry)  
 XX DE S. mutans glucosyltransferase GTFB mutant D457N/D567T.  
 XX KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX KM amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX OS Streptococcus mutans.  
 XX OS Synthetic.  
 XX PH Key Location/Qualifiers  
 XX FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"  
 XX FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
 XX FT US2002031826-A1.  
 XX PD 14-MAR-2002.  
 XX PF 19-DEC-2000; 2000US-00740274.  
 XX PR 07-JUN-1995; 95US-00478704.  
 XX PR 07-JUN-1995; 95US-00482711.  
 XX PR 07-JUN-1995; 95US-00485243.  
 XX PR 16-JAN-1998; 98US-00007999.  
 XX PR 16-JAN-1998; 98US-00008172.  
 XX PR 20-JAN-1998; 98US-00008620.  
 XX PR 11-DEC-1998; 98US-00210361.  
 XX PA (NICH/) NICHOLS S E.  
 XX PI Nichols SE;  
 XX DR WPI; 2002-414332/44.  
 XX PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 XX PT as substitutes for and additions to modified starch and latexes in paper  
 XX PT manufacture, comprises mutations in specific positions.  
 XX PS Claim 36; Page; 44p; English.  
 XX XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide, host  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing the glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 XX Query Match 81.8%; Score 90; DB 5; Length 1475;  
 XX Best Local Similarity 81.8%; Pred. No. 4.5e-06;  
 XX Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 XX DB 1 VPSYSFARAHDSVQDTIIRDI 22  
 XX 552 VPSYSFIRAHDSVQDTIADII 573  
 XX RESULT 17  
 XX AAU98032  
 XX ID AAU98032 standard; protein, 1475 AA.  
 XX AC AAU98032;  
 XX DT 27-AUG-2002 (first entry)  
 XX DE S. mutans glucosyltransferase GTFB mutant D567T.  
 XX KM Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX KM coating composition; glucan; starch; latex; thermoplastic molecule;  
 XX KM amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX OS Streptococcus mutans.  
 XX OS Synthetic.  
 XX PH Key Location/Qualifiers  
 XX FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
 XX FT US2002031826-A1.  
 XX PD 14-MAR-2002.  
 XX PF 19-DEC-2000; 2000US-00740274.  
 XX PR 07-JUN-1995; 95US-00478704.  
 XX PR 07-JUN-1995; 95US-00482711.  
 XX PR 07-JUN-1995; 95US-00485243.  
 XX PR 16-JAN-1998; 98US-00007999.  
 XX PR 16-JAN-1998; 98US-00008172.  
 XX PR 20-JAN-1998; 98US-00008620.  
 XX PR 11-DEC-1998; 98US-00210361.  
 XX PA (NICH/) NICHOLS S E.  
 XX PI Nichols SE;  
 XX DR WPI; 2002-414332/44.  
 XX PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 XX PT as substitutes for and additions to modified starch and latexes in paper  
 XX PT manufacture, comprises mutations in specific positions.  
 XX PS Claim 36; Page; 44p; English.  
 XX XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,



CC 1448b/D457M/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GFP D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GFP mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GFP mutant, an expression cassette comprising the expression cassette, host  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GFP, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll a/b binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GFP, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GFP mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GFPB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX SQ Sequence 1475 AA;

Query Match Best Local Similarity 81.8%; Score 90; DB 5; Length 1475;  
 Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSTSPARADSEVQDIIRDI 22  
 Db 552 VPSTSPARADSEVQDIIRDI 573

RESULT 18  
 ABR63235  
 ID ABR63235 standard; protein; 2055 AA.

XX ABR63235;

XX 23-OCT-2003 (revised)  
 DT 27-AUG-2003 (first entry)

XX Glucansucrase sequence from strain KG15.

XX Glucan: glucosyltransferase activity; thickener; prebiotic;  
 KM bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

XX Lactobacillus sakei.

XX WO2003008618-A2.

XX 30-JAN-2003.

XX 22-JUL-2002; 2002WO-NL000495.

XX 20-JUL-2001; 2001EP-00202752.  
 PR 25-JUL-2001; 2001EP-00202841.

XX (NEDE ) NEDERLANDSE ORG TOEGEPAST.

XX Van Geel- Schutten GH;  
 PI WPI; 2003-289780/28.  
 DR N-PSDB; ACC84451.  
 XX Novel glucan produced by glucosyltransferase activity of lactic acid  
 PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.  
 PS Claim 11; Fig 1; 51pp; English.

CC The present invention relates to glucan capable of being produced by  
 CC glucosyltransferase activity of a lactic acid bacterium on a sucrose  
 CC substrate. The method is useful as a thickener, as a prebiotic and as a  
 CC bioactive agent and as an anti-corrosion agent. The glucan can be  
 CC incorporated in foodstuffs such as beverages, sauces, dressings, dairy  
 CC products. The glucan is useful as anticorrosion agent, e.g. for the  
 CC protection of ship hulls. It can also be incorporated in nutritional or  
 CC pharmaceutical compositions intended for improving the condition of the  
 CC gastrointestinal tract. The present sequence represents a sequence of the  
 CC glucansucrase gene. (Updated on 23-OCT-2003 to standardise OS field)

XX SQ Sequence 2055 AA;

Query Match Best Local Similarity 78.2%; Score 86; DB 6; Length 2055;  
 Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSTSPARADSEVQDIIRDI 22  
 Db 955 IPSTSPARADSEVQDIIRDI 976

RESULT 19  
 ABR63236  
 ID ABR63236 standard; protein; 1149 AA.

XX ABR63236;

XX 27-AUG-2003 (first entry)

XX Glucansucrase sequence from strain lb33.

XX Glucan: glucosyltransferase activity; thickener; prebiotic;  
 KM bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.

XX Lactobacillus fermentum.

XX WO2003008618-A2.

XX 30-JAN-2003.

XX 22-JUL-2002; 2002WO-NL000495.

XX 20-JUL-2001; 2001EP-00202752.

XX 25-JUL-2001; 2001EP-00202841.

XX (NEDE ) NEDERLANDSE ORG TOEGEPAST.

XX Van Geel- Schutten GH;

XX WPI; 2003-289780/28.

XX N-PSDB; ACC84452.

XX Novel glucan produced by glucosyltransferase activity of lactic acid  
 PT bacterium on sucrose substrate, and having backbone consisting of alpha  
 PT (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.

XX Claim 11; Fig 1; 51pp; English.

CC The present invention relates to glucan capable of being produced by  
 CC glucosyltransferase activity of a lactic acid bacterium on a sucrose



PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX  
 PI Nichols SE;  
 XX  
 DR WPI; 2002-414332/44.  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Claim 36; Page; 44pp; English.  
 XX  
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing the glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilizing the glucan produced by GTF, which utilizes  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AU98027 and the information in  
 CC claim 36  
 CC  
 SQ Sequence 1475 AA;  
 Query Match 75.5%; Score 83; DB 5; Length 1475;  
 Best Local Similarity 77.3%; Pred. No. 7.2e-05;  
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 VPSYSPARAHSEVODIIRDI 22  
 DB 552 VPSYSPIRAHSEVOTLAKIT 573

RESULT 22  
 ID AAU98038 standard; protein; 1475 AA.  
 XX  
 AC AAU98038;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans GTFB mutant I448V/D457N/D567T/D571K/K779Q/K1014T.  
 XX  
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; muten.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX  
 XX  
 XX Key Location/Qualifiers  
 FH MISC-difference 448 /note= "Wild-type Ile substituted by Val"  
 FT MISC-difference 457 /note= "Wild-type Asp substituted by Asn"  
 FT MISC-difference 567 /note= "Wild-type Asp substituted by Thr"  
 FT MISC-difference 571 /note= "Wild-type Asp substituted by Lys"  
 FT MISC-difference 779 /note= "Wild-type Lys substituted by Gln"  
 FT MISC-difference 1014 /note= "Wild-type Lys substituted by Thr"  
 XX  
 XX US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX  
 PI Nichols SE;  
 XX  
 DR WPI; 2002-414332/44.  
 XX  
 PT Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Claim 36; Page; 44pp; English.  
 XX  
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from 1448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where



QY 2 PYSFARADSEVQDIIRDI 22  
 ID AUB80055 standard; protein, 1527 AA.  
 DB 625 PMSFVRADSEVQTVIAQIV 645

RESULT 25  
 ID AUB80055 standard; protein, 1527 AA.  
 DB 625 PMSFVRADSEVQTVIAQIV 645

AC AUB80055;

DT 30-JUL-2002 (first entry)

DE Leuconostoc mesenteroides dextranucrase.

KM Dextranucrase; yoghurt; curd; cheese; fermented milk; infant formulae;  
 KW pet food; vitamin; oral vaccine; enzyme.

OS Leuconostoc mesenteroides.

XX EP1201131-A1.

XX 02-MAY-2002.

PF 23-OCT-2000; 2000EP-00123012.

XX 23-OCT-2000; 2000EP-00123012.

PR (NEST ) SOC PROD NESTLE SA.

XX Bauche A, De Maleprade D, Duboc P, Neubauer H, Zink R;

XX WPI; 2002-373873/41.

DR N-PSDB; ABK50932.

PT Preparing fermented food products, e.g. yoghurt, using microorganisms  
 PT that have dextran sucrose bound to the surface, provide attractive and  
 uniform texture.

PS Disclosure; Page 14-20; 38pp; English.

XX The invention relates to preparation of a fermented food product that  
 CC includes treating the microorganism (A) to be used for fermentation with  
 CC a dextranucrase so that this binds to the cell walls of (A). The method  
 CC is used to produce yoghurt, curd, cheese or other fermented milk  
 CC products, ice cream, fermented cereal products, water-based jellies,  
 CC infant formulae and pet foods. Dextranucrase can also be bound to e.g.  
 CC vitamins or oral vaccines for delivering these to foods, using (A) as  
 CC carrier. The method produces foods with attractive and uniform texture,  
 CC since dextranucrase is evenly distributed and forms a thickening agent  
 CC in situ, with the amount formed being controlled by the amount of  
 CC dextranucrase bound. Milk does not have to be coagulated in order to  
 CC achieve a good texture, so the amount of acidic by-products formed is  
 CC reduced. Dextranucrase binds to many different types of cells over wide  
 CC ranges of temperature and pH and the treated bacteria can be stored in  
 CC liquid or powdered form. The present sequence represents the amino acid  
 CC sequence of dextranucrase

XX Sequence 1527 AA;

XX Query Match 73.6%; Score 81; DB 5; Length 1527;

XX Best Local Similarity 63.6%; Pred. No. 0.00017;  
 XX Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 PYSFARADSEVQDIIRDI 22

DB 652 IPNYSFVRADSEVQTVIAQIV 673

RESULT 26

ADCS4807 standard; protein, 1527 AA.

XX ADCS4807;

XX 18-DEC-2003 (first entry)

DE Leuconostoc mesenteroides dextran sucrose protein.

XX dextran sucrose; active centre zone; glucan; polysaccharide; dextran;  
 KW D-glucose; starch; cellulose; glucan manufacture; transduction;  
 KW enzyme-reaction product.

OS Leuconostoc mesenteroides.

XX JP2003111590-A.

XX 15-APR-2003.

XX 03-OCT-2001; 2001JP-00307067.

XX 03-OCT-2001; 2001JP-00307067.

PR (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUHIN SOGO KEN.

XX WPI; 2003-735670/70.

DR N-PSDB; ADCS4802.

PT Novel modified dextran sucrose which exchanges one site of active center  
 PT zone of dextran sucrose for active center zone of different types of  
 PT dextran sucrose, useful for manufacturing glucan.

PS Example 1; SEQ ID NO 12; 28pp; Japanese.

XX This invention relates to a modified dextran sucrose (DS) exchanging one  
 CC site of the active centre zone of a dextran sucrose for the active centre  
 CC zone of different types of dextran sucrose. Glucan is a polysaccharide  
 CC (for example dextran) which uses D-glucoses, such as a starch and a  
 CC cellulose, as a structural unit. The modified enzyme of the invention is  
 CC useful in the manufacture of glucan. The selection of the active centre  
 CC area of DS which carries out transduction, enables changes in structure  
 CC and character of an enzyme-reaction product and their application to  
 CC various uses. The present sequence is that of the Leuconostoc  
 CC mesenteroides dextran sucrose protein used during the exemplification of  
 CC the invention.

XX Sequence 1527 AA;

XX Query Match 73.6%; Score 81; DB 7; Length 1527;

XX Best Local Similarity 63.6%; Pred. No. 0.00017;  
 XX Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 PYSFARADSEVQDIIRDI 22

DB 652 IPNYSFVRADSEVQTVIAQIV 673

RESULT 27

ADCS4807 standard; protein, 1554 AA.

XX ADD93658;

XX 29-JAN-2004 (first entry)

DE Streptococcus sobrinus glucosyltransferase-U.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus sobrinus.

XX WO2003075845-A2.

XX 18-SEP-2003.



CC glucansucrase gene  
XX  
SQ Sequence 224 AA;

XX Sequence 224 AA;  
SQ

XX Sequence 224 AA;  
SQ

Query Match	69.1%	Score 76	DB 6	Length 224
Best Local Similarity	71.4%	Pred. No. 0.00012		
Matches 15, Conservative		2	Mismatches 4	Indels 0
				Gaps 0

```
QY      2 PSYSFARAHDSVQDIIRDII 22
        |:|:| | | | | | | | | |
Db     142 PNYTFIRAHDSVQTTIAQTI 162
```

RESULT 30  
APR 63 224

ID ABR63234 standard; protein; 1497 AA.

AC ABR63234;

DT 27-AUG-2003 (first entry)

DE GLucanucrase sequence from strain Lb33.

KW Glucan; glucosyltransferase activity; thickener; prebiotic;

KW bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract

OS Lactobacillus sp.

PN WO2003008618-A2.

PD 30-JAN-2003 :

PF 22-JUL-2002; 2002WO-NL000495.

PR 20-JUL-2001; 2001EP-00202752.

XX

[illegible]

XX  
WPT: 3003-389780/28

DR N-PSDB; ACC84450.  
XX

PT Novel glucan prod  
PT bacterium on sucr

PT (1,3) - and alpha (1,6) -linked dextran, used as crosslinker

XX PS CLAIM 11; FIG 1; SUPP; ENGLISH.

The present invention relates to glucan capable of being produced by glucosyltransferase activity of a lactic acid bacterium on a sucrose substrate. The method is useful as a thickener, as a prebiotic and as a bioactive agent and as an anti-corrosion agent. The glucan can be incorporated in foodstuffs such as beverages, sauces, dressings, dairy products. The glucan is useful as anticorrosion agent, e.g. for the protection of ship hulls. It can also be incorporated in nutritional or pharmaceutical compositions intended for improving the condition of the gastrointestinal tract. The present sequence represents a sequence of the glucanucrase gene

SQ Sequence 1497 AA;

Query Match	69.1%	Score 76;	DB 6;	Length 1497;
Best Local Similarity	71.4%	Pred. No. 0.0012;		
Matches 15;	Conservative	2;	Mismatches 4;	Indels 0; Gaps 0

```
QY      2 P5SYFARAHDSVQDIIRDII 22
        |::| | | | | | | | | |
Db      605 PNYYFIRAHDSVQTIIRAQII 625
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using SW model

Run on: May 4, 2004, 09:06:31 ; Search time 11.6667 Seconds  
(without alignments)  
181.390 Million cell updates/sec

Title: US-09-290-049a-19

Perfect score: 110  
Sequence: 1 VPSYSFARAHSEVQDIIRDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 100 summaries

## Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	110	100.0	A38175	glucosyltransferase
2	98	89.1	JT0345	dextranase (EC
3	97	88.2	B33135	glfB protein precu
4	85	77.3	UC5473	dextranase (EC
5	82	74.5	1508	probable dextran
6	72	65.5	S22737	glucosyltransferase
7	70	63.6	T30857	glucosyltransferase
8	69	62.7	T30552	glucosyltransferase
9	67	60.9	A44811	glucosyltransferase
10	64	58.2	A45866	dextranase (EC
11	64	58.2	T30858	glucosyltransferase
12	64	58.2	T30858	glucosyltransferase
13	51	46.4	S75540	galactose-1-phosph
14	49	44.5	E83788	hypothetical prote
15	46	41.8	T47184	hypothetical prote
16	46	41.8	OTHUCR	natriuretic peptid
17	46	41.8	6359	hypothetical prote
18	45.5	41.4	484	hypothetical prote
19	45	40.9	597	DNA polymerase III
20	45	40.9	597	DNA polymerase III
21	45	40.9	597	DNA polymerase III
22	44.5	40.5	486	glutamate-tRNA lig
23	44.5	40.5	603	glutamate-tRNA lig
24	44.5	40.5	603	glutamate-tRNA lig
25	44	40.0	188	cell fusion protei
26	44	40.0	216	L-fucose-6-phosph
27	44	40.0	330	atrazine chlorohyd
28	44	40.0	1058	retrovirus-related
29	43	39.1	68	hypothetical prote

30	43	39.1	176	F83819	hypothetical prote
31	43	39.1	295	A64236	hypothetical prote
32	43	39.1	419	A40728	microphthalmia-ass
33	43	39.1	475	AB2767	magnesium transpor
34	43	39.1	501	E97547	magnesium transpor
35	43	39.1	548	1 UFECAQ	glutamate hydratase
36	43	39.1	548	2 F90918	glutamate hydratase
37	43	39.1	548	2 F85767	glutamate hydratase
38	43	39.1	548	2 AB0691	glutamate hydratase
39	43	39.1	571	2 T08930	glutamate hydratase
40	42.5	38.6	395	2 B75047	glutamate hydratase
41	42.5	38.6	484	2 S73490	glutamate hydratase
42	42.5	38.6	540	2 S63299	sugar transport pr
43	42.5	38.6	688	2 S55349	potassium channel
44	42	38.2	196	2 S48302	MAD2 protein - Yea
45	42	38.2	217	2 T11382	cytochrome-c oxida
46	42	38.2	327	2 H59094	hypothetical prote
47	42	38.2	358	2 D96722	atrial natriuretic
48	42	38.2	536	2 A45409	natriuretic peptid
49	42	38.2	537	1 A28111	hypothetical prote
50	42	38.2	651	2 T16450	probable coiled-co
51	42	38.2	1115	2 T41342	perilaxin - rat
52	42	38.2	1389	2 I58157	hypothetical prote
53	41	37.3	195	2 F72718	hypothetical prote
54	41	37.3	284	2 C64158	conserved hypochet
55	41	37.3	336	2 AB0761	hypothetical prote
56	41	37.3	337	2 AB0761	hypothetical prote
57	41	37.3	344	2 T05104	hypothetical prote
58	41	37.3	361	2 AB7656	probable cell divi
59	41	37.3	414	2 B71125	hypothetical prote
60	41	37.3	444	2 F86329	nucleolar protein
61	41	37.3	508	2 T50180	proline-tRNA synth
62	41	37.3	570	2 F70332	DNA repair helicase
63	41	37.3	637	2 E75044	hypothetical prote
64	41	37.3	691	2 S48390	hypothetical prote
65	41	37.3	734	2 A82047	primosomal replica
66	41	37.3	824	1 S50767	S-receptor kinase
67	41	37.3	1684	2 S10789	amylase A-180 - al
68	40.5	36.8	245	2 AG1524	riboflavin kinase
69	40.5	36.8	246	2 AH1165	ABC transporter, P
70	40.5	36.8	287	2 H90462	ethylene-inducible
71	40.5	36.8	338	2 G90203	glutamate-tRNA lig
72	40.5	36.8	483	2 E86881	glutamate-tRNA lig
73	40.5	36.8	486	2 C95242	gp68.1 protein - M
74	40	36.4	80	2 D72808	hypothetical prote
75	40	36.4	100	2 D96005	hypothetical prote
76	40	36.4	169	2 S37892	hypothetical prote
77	40	36.4	194	2 A12364	hypothetical prote
78	40	36.4	205	2 C75155	methylenetetrahydr
79	40	36.4	207	2 C72223	guanylate kinase -
80	40	36.4	211	2 T41965	hypothetical prote
81	40	36.4	222	2 T09511	hypothetical prote
82	40	36.4	229	2 A11755	hypothetical prote
83	40	36.4	314	1 ITVZRX	bacteriophage prot
84	40	36.4	344	2 JH0511	poxvirus DNA topoi
85	40	36.4	359	2 G82197	inositol 2-dehydro
86	40	36.4	385	2 S64770	RetA1/RscA2 protei
87	40	36.4	429	2 S20050	hypothetical prote
88	40	36.4	451	2 T41129	transcription fact
89	40	36.4	491	2 AF1104	hypothetical prote
90	40	36.4	491	2 AF1104	hypothetical prote
91	40	36.4	495	2 A70782	glutamy1-tRNA synt
92	40	36.4	495	2 A70782	glutamy1-tRNA synt
93	40	36.4	575	2 S46329	probable monooxyge
94	40	36.4	635	2 T19654	intermediate filam
95	40	36.4	659	2 A64139	hypothetical prote
96	40	36.4	751	2 D71860	ribonuclease T2 ho
97	40	36.4	778	2 H95046	probable outer mem
98	40	36.4	788	2 H97917	MutS2 family prote
99	40	36.4	822	2 F69164	DNA mismatch repai
100	40	36.4	892	2 AB7447	ATP-dependent RNA



## ALIGNMENTS

## RESULT 1

A38175

glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C&gt;Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999

C/Accession: A38175

R:Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A&gt;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A/Reference number: A38175; MUID:91123227; PMID:1704006

A/Accession: A38175

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1592 &lt;ABO&gt;

A/Cross-references: GB:D90213; NID:G217032; PIDN:BA14241.1; PID:G217033

C/Superfamily: cpl repeat homology

F:1093-1112/Domain: cpl repeat homology &lt;CP1&gt;

F:1222-1241/Domain: cpl repeat homology &lt;CP2&gt;

F:1287-1306/Domain: cpl repeat homology &lt;CP3&gt;

F:1330-1351/Domain: cpl repeat homology &lt;CP4&gt;

F:1352-1371/Domain: cpl repeat homology &lt;CP5&gt;

F:1402-1420/Domain: cpl repeat homology &lt;CP6&gt;

F:1465-1484/Domain: cpl repeat homology &lt;CP7&gt;

F:1513-1532/Domain: cpl repeat homology &lt;CP8&gt;

Query Match

Best Local Similarity 100.0%; Score 110; DB 2; Length 1592;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSTFARAHDSVQDIIIRDII 22

Db 548 VPSTFARAHDSVQDIIIRDII 569

## RESULT 2

JTC0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N:Alternate names: sucrose 6-glucosyltransferase

C:Species: Streptococcus mutans

C&gt;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999

C/Accession: JTC0345; C33135

R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A&gt;Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A/Reference number: JTC0345; MUID:89137980; PMID:2976010

A/Accession: JTC0345

A/Molecule type: DNA

A/Residues: 1-1375 &lt;UED&gt;

A/Experimental source: GS-5

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A&gt;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A/Reference number: A33135; MUID:87308013; PMID:3040685

A/Accession: C33135

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-349 &lt;SHI&gt;

A/Cross-references: GB:M17361

C/Genetics:

A/Gene: gtfC

C/Function: catalyzes the synthesis of both water-soluble and water-insoluble glucans

C/Superfamily: cpl repeat homology

F:1-34/Domain: signal sequence #status predicted &lt;SIG&gt;

F:33-1375/Product: glucosyltransferase #status predicted &lt;MAT&gt;

F:1126-1145/Domain: cpl repeat homology &lt;CP1&gt;

F:1253-1272/Domain: cpl repeat homology &lt;CP2&gt;

F:1318-1337/Domain: cpl repeat homology &lt;CP3&gt;

Query Match 89.1%; Score 98; DB 2; Length 1375;

Best Local Similarity 86.4%; Pred. No. 4.5e-07; Mismatches 1; Indels 0; Gaps 0;

Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSTFARAHDSVQDIIIRDII 22

Db 578 VPSTFARAHDSVQDIIIRDII 599

## RESULT 3

B33135

gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C&gt;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

C/Accession: B33135; A33128

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A&gt;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A/Reference number: A33135; MUID:87308013; PMID:3040685

A/Accession: B33135

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1475 &lt;SHI&gt;

A/Cross-references: GB:M17361; NID:G153639; PIDN:AAA88588.1; PID:G153640

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

A/Reference number: A33128

A/Accession: A33128

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-171,173-641,'N',643-1475 &lt;SH2&gt;

A/Experimental source: strain GS-5

C/Superfamily: cpl repeat homology

F:1096-1115/Domain: cpl repeat homology &lt;CP1&gt;

F:1224-1243/Domain: cpl repeat homology &lt;CP2&gt;

F:1289-1308/Domain: cpl repeat homology &lt;CP3&gt;

F:1354-1373/Domain: cpl repeat homology &lt;CP4&gt;

F:1419-1438/Domain: cpl repeat homology &lt;CP5&gt;

Query Match

Best Local Similarity 88.2%; Score 97; DB 2; Length 1475;

Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSTFARAHDSVQDIIIRDII 22

Db 552 VPSTFARAHDSVQDIIIRDII 573

## RESULT 4

JCS473

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C&gt;Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

C/Accession: JCS473

R:Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monnan, P.

Gene 182, 23-32, 1996

A&gt;Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leucon

A/Reference number: JCS473; MUID:97136886; PMID:8982063

A/Accession: JCS473

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-1290 &lt;MON&gt;

A/Cross-references: GB:U38181

C/Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont

C/Genetics:

A/Gene: dera

C/Keywords: glycosyltransferase; hexosyltransferase

F:78-870/Domain: catalytic #status predicted &lt;CAT&gt;

F:922-1290/Domain: glucan-binding #status predicted &lt;GCB&gt;

Query Match

Best Local Similarity 77.3%; Score 85; DB 2; Length 1290;

Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PSYSPARAHDSVQDIIRDII 22  
 DB 388 NPYSPFRAHDSVQDIIRDII 408

## RESULT 5

T31098  
 probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
 C:Species: Leuconostoc mesenteroides  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T31098  
 R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
 FEBS Microbiol. Lett. 159, 307-315, 1998  
 A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS  
 A:Reference number: Z20981; MUID:98164374; PMID:9503626  
 C:Accession: T31098  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1508 <MON>  
 A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AA095453.1  
 A:Experimental source: strain NRRL B-1299  
 C:Genetics:  
 A:Gene: dsrB  
 C:Function:  
 A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 74.5%; Score 82; DB 2; Length 1508;  
 Best Local Similarity 68.2%; Pred. No. 0.00018;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSVQDIIRDII 22  
 DB 634 IPNYSPFRAHDSVQDIIRDII 655

## RESULT 6

S22737  
 glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
 C:Species: Streptococcus salivarius  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
 C:Accession: S22737; S28810; B44811; S22727  
 R:Jacques, N.  
 submitted to the EMBL Data Library, March 1992  
 A:Reference number: S22726  
 A:Accession: S22737  
 A:Molecule type: DNA  
 A:Residues: 1-1599 <JAC>  
 A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CA077898.1; PID:g47531  
 A:Experimental source: ATCC 25975  
 R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
 J. Gen. Microbiol. 137, 2577-2593, 1991  
 A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge  
 A:Reference number: A44811; MUID:92148377; PMID:1838391  
 A:Accession: S28810  
 A:Molecule type: DNA  
 A:Residues: 1-51 <GIF>  
 A:Cross-references: EMBL:Z11873  
 C:Genetics:  
 A:Gene: gtfK  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F:156-1475/Domain: cpl repeat homology <CPR>

Query Match 65.5%; Score 72; DB 2; Length 1599;  
 Best Local Similarity 70.0%; Pred. No. 0.0073;  
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHDSVQDIIRDII 22  
 DB 574 TYLFVRAHDSVQDIIRDII 593

## RESULT 7

T30857  
 glucosyltransferase - Streptococcus salivarius  
 C:Species: Streptococcus salivarius  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30857  
 R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
 Infect. Immun. 63, 609-621, 1995  
 A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for p  
 A:Reference number: Z20909; MUID:95122197; PMID:7822030  
 A:Accession: T30857  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1449 <SIM>  
 A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AA041412.1  
 C:Genetics:  
 A:Gene: gtfL

Query Match 63.6%; Score 70; DB 2; Length 1449;  
 Best Local Similarity 70.0%; Pred. No. 0.014;  
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHDSVQDIIRDII 22  
 DB 609 NYAFVRAHDSVQDIIRDII 628

## RESULT 8

T30552  
 glucosyltransferase N - Streptococcus salivarius (fragment)  
 C:Species: Streptococcus salivarius  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30552  
 R:Jaife, R.I.  
 submitted to the EMBL Data Library, February 1998  
 A:Description: Streptococcus salivarius V1477 gtfN.  
 A:Reference number: Z20854  
 A:Accession: T30552  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1449 <JAF>  
 A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AA05156.1  
 C:Genetics:  
 A:Gene: gtfN

Query Match 63.6%; Score 70; DB 2; Length 1449;  
 Best Local Similarity 70.0%; Pred. No. 0.014;  
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHDSVQDIIRDII 22  
 DB 609 NYAFVRAHDSVQDIIRDII 628

## RESULT 9

A41483  
 glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
 C:Species: Streptococcus sobrinus  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
 C:Accession: A41483  
 R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
 Infect. Immun. 58, 2452-2458, 1990  
 A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltransf  
 A:Reference number: A41483; MUID:90316665; PMID:2124279  
 A:Accession: A41483  
 A:Molecule type: DNA  
 A:Residues: 1-1365 <GIL>  
 A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653  
 C:Genetics:  
 A:Gene: gtfS  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 62.7%; Score 69; DB 2; Length 1365;  
Best Local Similarity 68.2%; Pred. No. 0.018;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSFPAHDSVQDIIRDI 22  
DB 537 VPIVFAHDSVQVIRAKII 558

## RESULT 10

A:Accession: A44811  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377; PMID:1838391  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GIF>  
A:Cross-references: EMBL:Z11873; NID:G47526; PDB:CAA72900.1; PID:G47527  
A:Note: sequence extracted from NCBI Backbone (NCBIN:81050, NCBI:81052)  
A:Genetics: gtfm  
A:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 60.3%; Score 67; DB 2; Length 1518;  
Best Local Similarity 65.0%; Pred. No. 0.043;  
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPAHDSVQDIIRDI 22  
DB 604 NYVFAHDSVQVIRAKII 623

## RESULT 11

A:Accession: A45866  
A:Title: dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: A45866  
A:Residues: 1-108 <FAN>  
A:Cross-references: EMBL:D90911; GB:AB001339; NID:G1653083; PDB:BA19101.1; PID:G165314  
A:Note: the nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl  
A:Reference number: A45866; MUID:91100958; PMID:2148600  
A:Accession: A45866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
A:Cross-references: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP6>  
F:1341-1361/Domain: cpl repeat homology <CP6>  
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 58.2%; Score 64; DB 2; Length 1431;  
Best Local Similarity 65.0%; Pred. No. 0.12;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSPAHDSVQDIIRDI 22  
DB 576 NYVFAHDSVQVIRAKII 595

RESULT 12  
T30858  
glucosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:G662380; PID:G662381; PDB:AAQ41413.1  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:G662380; PID:G662381; PDB:AAQ41413.1  
A:Genetics: gtfm

Query Match 58.2%; Score 64; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.13;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSPAHDSVQDIIRDI 22  
DB 661 NYVFAHDSVQVIRAKII 680

## RESULT 13

A:Accession: S75540  
A:Title: hypothetical protein sll1219 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S75540  
A:Residues: 1-108 <FAN>  
A:Cross-references: EMBL:D90911; GB:AB001339; NID:G1653083; PDB:BA19101.1; PID:G165314  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
A:Superfamily: Synechocystis hypothetical protein sll1219  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S75540  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-108 <FAN>  
A:Cross-references: EMBL:D90911; GB:AB001339; NID:G1653083; PDB:BA19101.1; PID:G165314  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
A:Superfamily: Synechocystis hypothetical protein sll1219

Query Match 46.4%; Score 51; DB 2; Length 108;  
Best Local Similarity 38.9%; Pred. No. 0.72;  
Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 5 SPARHDSVQDIIRDI 22  
DB 34 NMAHDSVQVIRAKII 51

## RESULT 14

A:Accession: E83788  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: E83788  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <STO>

A;Cross-references: GB:AP001510; GB:BA000004; NID:910173440; PIDN:BA04828.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: galP  
C;Superfamily: galactose-1-phosphate uridylyltransferase, Bacillus type

Query Match 44.8%; Score 49; DB 2; Length 508;  
Best Local Similarity 62.5%; Pred. No. 8.9;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 7 ARADSEVODIIRDI 22  
Db 460 AHLDNSVMDLRDI 475

RESULT 15  
T47184  
hypothetical protein DKFZp434F1526.1 - human (fragment)

C;Species: Homo sapiens (man)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C;Accession: T47184  
R;Annotator: W. J. Mikner, U. J. Mewes, H. W. J. Well, B. J. Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A;Reference number: Z24375

A;Accession: T47184  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-506 <AAA>  
A;Cross-references: EMBL:AL62051  
A;Experimental source: adult testis; clone DKFZp434F1526  
C;Genetics:  
A;Note: DKFZp434F1526.1

Query Match 41.8%; Score 46; DB 2; Length 506;  
Best Local Similarity 36.4%; Pred. No. 27;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VPSYFARADSEVODIIRDI 22  
Db 456 VPQYFNSLPPSEKDLIQYV 477

RESULT 16  
OYHUCR

natriuretic peptide receptor C precursor - human  
N;Alternate names: atrial natriuretic factor clearance receptor; natriuretic peptide cle

C;Species: Homo sapiens (man)  
C;Date: 09-Nov-1990 #sequence\_revision 05-May-1995 #text\_change 22-Jun-1995

C;Accession: S10150; A35896  
R;Lowe, D.G.; Camarero, T.R.; Goedel, D.V.

Nucleic Acids Res. 18, 3412, 1990  
A;Title: cDNA sequence of the human atrial natriuretic peptide clearance receptor.

A;Reference number: S10150; MUID:90287735; PMID:2165522

A;Accession: S10150  
A;Molecule type: mRNA

A;Residues: 1-540 <LOW>  
A;Cross-references: EMBL:X52282; NID:g28705; PIDN:CAA36523.1; PID:g28706

A;Note: alternative splice form C6  
R;Porter, J.G.; Arsten, A.; Fuller, F.; Miller, J.A.; Gregory, L.C.; Lewicki, J.A.

Biophys. Res. Commun. 171, 796-803, 1990  
A;Title: Isolation and functional expression of the human atrial natriuretic peptide cle

A;Reference number: A35896; MUID:90386656; PMID:2169733

A;Accession: A35896  
A;Molecule type: mRNA

A;Residues: 1-475, 'SG', 477-540 <POR>  
A;Cross-references: GB:M5305; NID:9178651; PIDN:AAA51734.1; PID:9178652

A;Note: alternative splice form C5  
R;Stults, J.T.; O'Connell, K.L.; Garcia, C.; Wong, S.; Engel, A.M.; Garbers, D.L.; Lowe,

Biochemistry 33, 11372-11381, 1994

A;Title: The disulfide linkages and glycosylation sites of the human natriuretic peptide

A;Reference number: A55870; MUID:9524450; PMID:7727388

A;Contents: annotation; propeptide, disulfide bonds, carbohydrate binding sites  
C;Genetics:

A;Gene: GDB:NPR3; NPRC; ANPRC  
A;Cross-references: GDB:125201; OMIM:108962

A;Map position: 5p14-5p13  
A;Superfamily: natriuretic peptide receptor C; natriuretic peptide-binding domain homolo

C;Keywords: alternative splicing; glycoprotein; homodimer; hormone receptor; transmembr

F;1-23/Domain: signal sequence #status predicted <SIG>

F;24-45/Domain: propeptide #status experimental <PRO>

F;46-50/Domain: natriuretic peptide receptor C #status predicted <MAT>

F;51-462/Domain: extracellular #status predicted <EXT>

F;483-503/Domain: transmembrane #status predicted <TM>

F;108-293/Binding site: carbohydrate (Asn) (covalent) #status experimental

F;394/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental

F;473-476/Disulfide bonds: (partial) (in splice form C6) #status experimental

Query Match 41.8%; Score 46; DB 1; Length 540;  
Best Local Similarity 44.4%; Pred. No. 29;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YSFARADSEVODIIRDI 21  
Db 234 YSFDEKDLDIRI 251

RESULT 17  
T31679

bactratin synthetase 3 - Bacillus licheniformis  
C;Species: Bacillus licheniformis

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
C;Accession: T31679

R;Korn, D.; Klenz, A.; Schorgerdorfer, K.; Marahiel, M.A.

Chem. Biol. 4, 927-937, 1997

A;Title: The bactratin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecu

A;Reference number: 221058; MUID:98089193; PMID:9427658

A;Accession: T31679  
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA  
A;Residues: 1-6359 <KON>

A;Cross-references: EMBL:AF007865; NID:g4464275; PID:g2982196; PIDN:AAC06348.1

C;Genetics:

C;Superfamily: acetate-CoA ligase homology; acyl carrier protein homology

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;505-946/Domain: acetate-CoA ligase homology <ACLI>

F;964-1032/Domain: acyl carrier protein homology <ACP1>

F;1542-1978/Domain: acetyl carrier protein homology <ACP2>

F;31046-3483/Domain: acyl carrier protein homology <ACP3>

F;3500-3568/Domain: acyl carrier protein homology <ACP4>

F;4078-4526/Domain: acetate-CoA ligase homology <ACLP>

F;4542-4609/Domain: acyl carrier protein homology <ACP5>

F;5593-6032/Domain: acetate-CoA ligase homology <ACLS>

F;6050-6118/Domain: acyl carrier protein homology <ACPS>

F;3532/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 41.8%; Score 46; DB 2; Length 6359;  
Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
Matches 11; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

Qy 1 VPSYFARADSEVODIIRDI 22  
Db 3186 VPSFSEF-----DSVVDIITTLI 3203

RESULT 18  
A89820

glutaryl-tRNA synthetase [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C;Accession: A89820

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;



AB1229  
excinnuclease ABC chain C homolog uvrc [imported] - *Listeria monocytogenes* (strain EGD-e)  
C/Species: *Listeria monocytogenes*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C/Accession: AB1229  
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A/Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of *Listeria* species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AB1229  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <GUA>  
A/Cross-references: GB:NC\_003210; PIDN:CA09312.1; PID:G16410650; GSPDB:GN00177  
C/Genetics:  
A/Experimental source: strain EGD-e  
C/Genetics:  
A/Gene: uvrc  
C/Superfamily: excinnuclease ABC chain C

Query Match 40.5%; Score 44.5; DB 2; Length 603;  
Best Local Similarity 40.9%; Pred. No. 56;  
Matches 9; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

Qy 1 VPSSFPARHDSVQDITRDI 22  
Db 42 VRSY-FSGTHDSKTRQLVQEV 62

RESULT 24  
AD1582  
excinnuclease ABC chain C homolog uvrc [imported] - *Listeria innocua* (strain C1p11262)  
C/Species: *Listeria innocua*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C/Accession: AD1582  
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.  
Science 294, 849-852, 2001  
A/Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of *Listeria* species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AD1582  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <GUA>  
A/Cross-references: GB:AL592022; PIDN:CA09312.1; PID:G16413656; GSPDB:GN00178  
A/Experimental source: strain C1p11262  
C/Genetics:  
A/Gene: uvrc  
C/Superfamily: excinnuclease ABC chain C

Query Match 40.5%; Score 44.5; DB 2; Length 603;  
Best Local Similarity 40.9%; Pred. No. 56;  
Matches 9; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

Qy 1 VPSSFPARHDSVQDITRDI 22  
Db 42 VRSY-FSGTHDSKTRQLVQEV 62

RESULT 25  
S31626  
cell fusion protein - myxoma virus  
C/Species: myxoma virus  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Nov-1999  
C/Accession: S31626  
R/Jackson, R.J.  
Submitted to the EMBL Data Library, January 1993

A/Description: A myxoma virus gene with similarity to the vaccinia virus gene encoding t  
A/Reference number: S31626  
A/Accession: S31626  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-188 <AAC>  
A/Cross-references: EMBL:Z19600; NID:G60609; PIDN:CA09312.1; PID:G60610  
C/Superfamily: vaccinia virus 14K cell fusion protein

Query Match 40.0%; Score 44; DB 2; Length 188;  
Best Local Similarity 35.3%; Pred. No. 18;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 4 YSPARHDSVQDITRDI 20  
Db 106 YNFAIQERDISDILKD 122

RESULT 26  
AF1336  
L-fuculose-phosphate aldolase homolog lmo2094 [imported] - *Listeria monocytogenes* (strai  
C/Species: *Listeria monocytogenes*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C/Accession: AF1336  
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.  
Science 294, 849-852, 2001  
A/Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of *Listeria* species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AF1336  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-216 <GUA>  
A/Cross-references: GB:NC\_003210; PIDN:CA00172.1; PID:G16411564; GSPDB:GN00177  
A/Experimental source: strain EGD-e  
C/Genetics:  
A/Gene: lmo2094

Query Match 40.0%; Score 44; DB 2; Length 216;  
Best Local Similarity 42.1%; Pred. No. 21;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPSSFPARHDSVQDITR 19  
Db 126 IPTAFAPATSSSLAIVR 144

RESULT 27  
F69471  
atrazine chlorohydrolase homolog - *Archaeoglobus fulgidus*  
C/Species: *Archaeoglobus fulgidus*  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C/Accession: F69471  
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
S.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A/Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Attlich, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Weese, C.R.; Venter, J.C.  
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A/Reference number: A69250; MUID:9804343; PMID:9389475  
A/Accession: F69471  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-330 <KLE>  
A/Cross-references: GB:AB000980; GB:AB000782; NID:G2689303; PIDN:AA089475.1; PID:G264877

Query Match 40.0%; Score 44; DB 2; Length 330;  
Best Local Similarity 42.1%; Pred. No. 34;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:02:41 ; Search time 8 Seconds  
(Without alignments)  
143.193 Million cell updates/sec

Title: US-09-290-049a-19  
Perfect score: 110  
Sequence: 1 VPSYSPARAHDSVQDIIRDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1592	1	PT7470 streptococ
2	108	98.2	1597	1	PT1001 streptococ
3	103	93.6	1476	1	PT8987 streptococ
4	98	89.1	1455	1	PT3470 streptococ
5	69	62.7	1365	1	PT9336 streptococ
6	64	58.2	1462	1	PT9331 streptococ
7	49	44.5	508	1	QKXV2 bacillus ha
8	46.5	42.3	484	1	Q8CTJ3 streptococ
9	46	41.8	541	1	PT7342 homo sapien
10	46	41.4	6359	1	Q68008 b bacitraci
11	44.5	40.5	484	1	Q99W75 streptococ
12	44.5	40.5	486	1	Q8CWM5 streptococ
13	44.5	40.5	603	1	Q22CH5 listeria in
14	44.5	40.5	330	1	Q8Y7P0 listeria mo
15	44	40.0	330	1	Q28499 archaeoglob
16	44	40.0	1046	1	P61777 weissella h
17	44	40.0	1058	1	P04323 diosiphilla
18	44	39.1	258	1	Q88GV1 rhizobium 1
19	43	39.1	295	1	P47566 mycoplasma
20	43	39.1	526	1	Q08874 mus musculu
21	43	39.1	547	1	P00923 escherichia
22	43	39.1	557	1	Q3WV0 arabidopsis
23	43	39.1	579	1	P40720 salmonella
24	43	39.1	799	1	Q86W18 mycoplasma
25	43	39.1	993	1	Q8Y7P0 listeria faa
26	42.5	38.6	395	1	Q9UYZ1 pyrococcus
27	42.5	38.6	484	1	P5114 mycoplasma
28	42.5	38.6	540	1	P42833 saccharomyc
29	42	38.2	196	1	P40958 saccharomyc
30	42	38.2	537	1	P10730 bos taurus
31	42	38.2	1383	1	Q63425 rattus norv
32	41.5	37.7	489	1	Q8E002 oceanobacil
33	41	37.3	109	1	Q9ELW1 cyanidium c

34	41	37.3	190	1	DCD_CHLVC
35	41	37.3	280	1	YIBQ_HAEIN
36	41	37.3	306	1	FMRE_LYMST
37	41	37.3	414	1	PT22_PYROHO
38	41	37.3	453	1	MDHP_FIABI
39	41	37.3	491	1	YIOA_YEAST
40	40.5	36.8	481	1	SYE_STRP3
41	40.5	36.8	481	1	SYE_STRP3
42	40.5	36.8	481	1	SYE_STRP3
43	40.5	36.8	481	1	SYE_STRP3
44	40.5	36.8	481	1	SYE_STRP3
45	40.5	36.8	481	1	SYE_STRP3
46	40.5	36.8	481	1	SYE_STRP3
47	40	36.4	169	1	YKHO_YEAST
48	40	36.4	207	1	KGUA_THEMA
49	40	36.4	211	1	UL97_HSV7J
50	40	36.4	221	1	RP9_HUMAN
51	40	36.4	314	1	TOP1_SFVKA
52	40	36.4	344	1	M12D_BACSU
53	40	36.4	373	1	CD62_METMA
54	40	36.4	429	1	TF3A_YEAST
55	40	36.4	485	1	SYE_BACCA
56	40	36.4	485	1	SYE_BACCA
57	40	36.4	491	1	SYE_LISIN
58	40	36.4	491	1	SYE_LISIN
59	40	36.4	495	1	Y892_MYCTU
60	40	36.4	568	1	TREA_XANCP
61	40	36.4	638	1	HTPG_STRCO
62	40	36.4	659	1	RNB_HARIN
63	40	36.4	660	1	PXN1_COREF
64	40	36.4	1018	1	VGMN_BPMV
65	40	36.4	1068	1	P11A_BOVIN
66	40	36.4	1068	1	P11A_HUMAN
67	40	36.4	1068	1	P11A_MOUSE
68	40	36.4	1461	1	PRAK_HUMAN
69	40	36.4	1835	1	DURI_YEAST
70	39.5	35.9	183	1	RETB_BOVIN
71	39.5	35.9	201	1	RETB_PIG
72	39.5	35.9	297	1	ARGB_ANASP
73	39.5	35.9	297	1	ARGB_SYNY3
74	39.5	35.9	473	1	SYE_WIGOR
75	39.5	35.9	505	1	YG20_METUA
76	39.5	35.5	96	1	IBBA_PEA
77	39.5	35.5	114	1	IBB2_PEA
78	39.5	35.5	124	1	RBS_PSEHY
79	39.5	35.5	129	1	Y207_AERPE
80	39.5	35.5	338	1	FEN_METMA
81	39.5	35.5	373	1	CD62_METMA
82	39.5	35.5	386	1	METL_DICCA
83	39.5	35.5	416	1	FCTA_ECO57
84	39.5	35.5	416	1	FCTA_ECO57
85	39.5	35.5	416	1	FCTA_ECO57
86	39.5	35.5	436	1	THIC_CLOPE
87	39.5	35.5	540	1	SCK_HUMAN
88	39.5	35.5	541	1	EHDA_HUMAN
89	39.5	35.5	541	1	EHDA_MOUSE
90	39.5	35.5	588	1	SYD_XANAC
91	39.5	35.5	588	1	SYD_XANAC
92	39.5	35.5	679	1	FTSH_BACPR
93	39.5	35.5	702	1	MADE_SCHPO
94	39.5	35.5	702	1	MADE_SCHPO
95	39.5	35.5	788	1	YAI4_YEAST
96	39.5	35.5	881	1	HIRI_YEAST
97	39.5	35.5	933	1	ODOL_ECOLI
98	39.5	35.5	953	1	COBP_HUMAN
99	39.5	35.5	953	1	COBP_MOUSE
100	39.5	35.5	953	1	COBP_RAT

## ALIGNMENTS

RESULT 1



GTF2\_STRDO  
ID GTF2\_STRDO STANDARD: PRT, 1592 AA.  
AC P27470;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucosyltransferase-1 precursor (BC 2.4.1.5) (GTF-1) (Dextranucrase)  
OS (Streptococcus downei (Streptococcus sobrinus)).  
OC Streptococcus; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=13317;  
OK NCB1  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6715 / Serotype G;  
RX MEDLINE=91123227; PubMed=1704006;  
RA Abo H., Matsunuma T., Kodama T., Onta H., Fukui K., Kato K.,  
RT Kagawa H.;  
RT "Peptide sequences for sucrose splitting and glucan binding within  
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
RT synthetase).";  
RL J. Bacteriol. 173:989-996(1991).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6)-alpha-D-glucosyl (N) = D-  
CC fructose + (1,6)-alpha-D-glucosyl (N+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S1 SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
CC -1- SIMILARITY: Contains 16 cell wall binding repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: D90213; BAA14241.1; -  
CC InterPro: IPR002479; CW binding.  
CC InterPro: IPR003318; Glyco\_hydro\_70.  
CC Pfam: PF01473; CW\_binding\_1; 13.  
CC Pfam: PF02324; Glyco\_hydro\_70; 1.  
CC Transferrase; Glucosyltransferase; Signal; Repeat; Dental carries.  
CC KW TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARRIES.  
CC FT SIGNAL 1 38 POTENTIAL.  
CC FT CHAIN 1 38 GLUCOSYLTRANSFERASE-1.  
FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
FT REPEAT 1093 1142 1.  
FT REPEAT 1158 1207 2.  
FT REPEAT 1222 1272 3.  
FT REPEAT 1287 1337 4.  
FT REPEAT 1402 1451 5.  
FT REPEAT 1514 1563 6.  
FT REPEAT 1577 1592 7 (INCOMPLETE).  
SQ SEQUENCE 1592 AA; 176167 MW; BCD066D079351BCF CRC64;  
Query Match 100.0%; Score 110; DB 1; Length 1592;  
Best Local Similarity 100.0%; Pred. No. 4.6e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2  
GTF1\_STRDO  
ID GTF1\_STRDO STANDARD: PRT, 1597 AA.  
AC P11001;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucosyltransferase-1 precursor (BC 2.4.1.5) (GTF-1) (Dextranucrase)  
OS (Streptococcus downei (Streptococcus sobrinus)).  
OC Streptococcus; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=13317;  
OK NCB1  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MFE28;  
RX MEDLINE=87308014; PubMed=3040686;  
RA Ferretti J.J., Gilpin M.L., Russel R.R.B.;  
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
RT sobrinus MFE28";  
RL J. Bacteriol. 169:4271-4278(1987).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6)-alpha-D-glucosyl (N) = D-  
CC fructose + (1,6)-alpha-D-glucosyl (N+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S1 SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
CC -1- SIMILARITY: Contains 19 cell wall binding repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: M17391; AAC63063.1; -  
CC InterPro: IPR002479; CW binding.  
CC InterPro: IPR003318; Glyco\_hydro\_70.  
CC Pfam: PF01473; CW\_binding\_1; 16.  
CC Pfam: PF02324; Glyco\_hydro\_70; 1.  
CC Transferrase; Glucosyltransferase; Signal; Repeat; Dental carries.  
CC KW TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARRIES.  
CC FT SIGNAL 1 38 POTENTIAL.  
CC FT CHAIN 1 38 GLUCOSYLTRANSFERASE-1.  
FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
FT REPEAT 1099 1132 A REPEAT.  
FT REPEAT 1163 1213 AC REPEAT.  
FT REPEAT 1227 1277 AC REPEAT.  
FT REPEAT 1292 1342 AC REPEAT.  
FT REPEAT 1352 1399 B REPEAT.  
FT REPEAT 1406 1455 AC REPEAT.  
FT REPEAT 1465 1512 B REPEAT.  
FT REPEAT 1519 1568 AC REPEAT.  
FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
SQ SEQUENCE 1597 AA; 177080 MW; B9B86A200868798E CRC64;  
Query Match 98.2%; Score 108; DB 1; Length 1597;  
Best Local Similarity 95.5%; Pred. No. 9.6e-09;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 3  
 GTFB\_STRMU STANDARD; PRT: 1476 AA.  
 AC P08987; 069381; 069384; 069387; 069390; 069396;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
 GN (Sucrose 6-glucosyltransferase).  
 OS GTFB OR SMU.1004.  
 OC Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=GS-5;  
 RC MEDLINE=87308013; PubMed=3040685;  
 RX Shirota T., Ueda S., Kuramitsu H.K.;  
 RA "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,  
 RC MT4467 / Serotype G, and MT8148 / Serotype C;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans."; J. Bacteriol. 161:331-336(1998).  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=UA159 / ATCC 700610 / Serotype C;  
 RC MEDLINE=22295063; PubMed=12397186;  
 RX Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 RT pathogen."; Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 RL [4]  
 RN FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl} (N) = D-  
 CC fructose + {(1,6)-alpha-D-glucosyl} (N+1).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6-LINKAGES). GTF-S SYNTHESIZES BOTH  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
 CC -1- SIMILARITY: Contains 10 cell wall binding repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL, M17361; AAA88588.1; -  
 CC EMBL, D88651; BAA26101.1; -  
 CC EMBL, D88654; BAA26105.1; -  
 CC EMBL, D88657; BAA26109.1; -  
 CC EMBL, D88660; BAA26113.1; -  
 CC EMBL, D89977; BAA26119.1; -  
 CC EMBL, AE014940; AAN58705.1; -  
 CC PIR, B3135; B3135.

DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 7.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries;  
 KW Complete proteome.  
 FT SIGNAL 1 34  
 FT CHAIN 35 1476  
 FT DOMAIN 35 1051  
 FT DOMAIN 1097 1476  
 FT REPEAT 1097 1130  
 FT DOMAIN 1161 1470  
 FT REPEAT 1161 1210  
 FT REPEAT 1225 1275  
 FT REPEAT 1290 1340  
 FT REPEAT 1355 1405  
 FT REPEAT 1420 1470  
 FT REPEAT 62 62  
 FT VARIANT 62 65  
 FT VARIANT 65 68  
 FT VARIANT 68 68  
 FT VARIANT 78 78  
 FT VARIANT 86 86  
 FT VARIANT 89 89  
 FT VARIANT 168 168  
 FT VARIANT 276 276  
 FT VARIANT 399 399  
 FT VARIANT 474 474  
 FT VARIANT 512 512  
 FT VARIANT 519 519  
 FT VARIANT 701 701  
 FT VARIANT 708 708  
 FT VARIANT 938 938  
 FT VARIANT 952 957  
 FT VARIANT 963 964  
 FT VARIANT 968 970  
 FT VARIANT 1086 1086  
 FT VARIANT 1158 1158  
 FT VARIANT 1163 1163  
 FT VARIANT 1168 1168  
 FT VARIANT 1182 1182  
 FT VARIANT 1234 1234  
 FT VARIANT 1263 1263  
 FT VARIANT 1263 1263  
 FT VARIANT 1264 1264  
 FT VARIANT 1272 1272  
 FT VARIANT 1329 1329  
 FT VARIANT 1394 1394  
 FT VARIANT 1402 1402  
 FT VARIANT 1459 1459  
 FT VARIANT 570 570  
 FT CONFLICT 800 817  
 FT CONFLICT 1310 1310  
 FT CONFLICT 1476 AA; 165846 MW; 90609F31B4C6F CRC64;  
 SQ SEQUENCE  
 Query Match 93.6%; Score 103; DB 1; Length 1476;  
 Best Local Similarity 90.9%; Pred. No. 5; 3e-08;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VPSYSPARAHSEVQDIIRDI 22  
 DB 552 VPSYSPARAHSEVQDIIRDI 573

RESULT 4  
ID GTF\_C STREMU STANDARD; PERT; 1455 AA.  
AC P13470; 069382; 069385; 069388; 069391; 069397; P05427;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update) (GTF-SI)  
DE Glucosyltransferase-St precursor (PC 2.4.1.5) (GTF-SI)  
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).  
GN GTF\_C OR SMU.1005.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=89137980; PubMed=2976010;  
RA Ueda S., Shirota T., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";  
RL Gene 69:101-109(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,  
RX MT4467 / Serotype B, and MT8148 / Serotype C;  
RA Fujikawa T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=UA159 / ATCC 700610 / Serotype C;  
RX MEDLINE=22295063; PubMed=12397186;  
RA Ajdic D., Meschan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
pathogen.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
RN [4]  
RP SEQUENCE OF 1-349 FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=87308013; PubMed=3040685;  
RA Shirota T., Ueda S., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
RL J. Bacteriol. 169:4263-4270(1987).  
RN [5]  
RP FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl) (N) = D-  
fructose + ((1,6)-alpha-D-glucosyl) (N+1).  
CC -1- SUBCELLULAR LOCATION: Secreted  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES BOTH  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE), GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
CC -1- SIMILARITY: Contains 5 cell wall binding repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D88652; BAA26102.1; -;  
DR EMBL; M22054; AAA88592.1; -;  
DR EMBL; D88652; BAA26102.1; -;

DR EMBL; D88655; BAA26106.1; -;  
DR EMBL; D88658; BAA26110.1; -;  
DR EMBL; D88661; BAA26114.1; -;  
DR EMBL; D89978; BAA26120.1; -;  
DR EMBL; AE014940; ANS6706.1; -;  
DR EMBL; M17361; AAA88589.1; -;  
DR PIR; J0345; J0345.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR003318; Glyco hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Pfam; PF02324; Glycosyltransferase; Signal; Repeat; Dental caries;  
KW Complete proteome.  
FT SIGNAL 1 34  
FT CHAIN 35 1455  
FT DOMAIN 35 1050  
FT DOMAIN 1126 1455  
FT DOMAIN 1126 1455  
FT REPEAT 1126 1159  
FT REPEAT 1169 1200  
FT REPEAT 1227 1238  
FT REPEAT 1253 1303  
FT REPEAT 1318 1330  
FT REPEAT 21 21  
FT VARIANT 81 81  
FT VARIANT 106 106  
FT VARIANT 116 116  
FT VARIANT 126 126  
FT VARIANT 150 151  
FT VARIANT 256 256  
FT VARIANT 425 425  
FT VARIANT 519 519  
FT VARIANT 538 538  
FT VARIANT 545 545  
FT VARIANT 597 597  
FT VARIANT 600 600  
FT VARIANT 601 601  
FT VARIANT 614 614  
FT VARIANT 727 727  
FT VARIANT 734 734  
FT VARIANT 964 964  
FT VARIANT 1113 1113  
FT VARIANT 1118 1118  
FT VARIANT 1204 1204  
FT VARIANT 1208 1208  
FT VARIANT 1292 1294  
FT VARIANT 1305 1369  
FT VARIANT 1326 1326  
FT VARIANT 1331 1331  
FT VARIANT 1377 1377  
FT VARIANT 1398 1398  
FT VARIANT 1424 1424  
FT VARIANT 1439 1439  
FT VARIANT 1444 1444  
FT VARIANT 1455 1455  
FT CONFLICT 1337 1455  
SO SEQUENCE 1455 AA; 162965 MW; 3CB455A9944FEC86 CRC64;  
Query Match 89.1%; Score 98; DB 1; Length 1455;  
Best local similarity 86.4%; Pred. No. 3.2e-07;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 VPSPARADSEVODIIRDI 22  
||||| ||||||| |||||

Db 578 VPSYSFIRAHDSVQDILIRMI 599

## RESULT 5

GTFS STRDO STANDARD; PRT; 1365 AA.  
AC P29336;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucosyltransferase-5 precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN GTF-S.  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MFE28;  
RX MEDLINE=9031665; PubMed=2142479;  
RA Gilmore K.S., Russell R.R., Perretti J.J.;  
RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
glucosyltransferase that synthesizes soluble glucans.";  
RL Infect. Immun. 58:2452-2458 (1990).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl) (N) = D-  
fructose + ((1,6)-alpha-D-glucosyl) (N+1).  
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
PRIMER GLUCAN UNLIKE GTF-1.  
CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
1,6-GLUCOSE).  
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
CC -1- SIMILARITY: Contains 10 cell wall binding repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M30943; AAA26898.1; -  
DR InterPro; IPR002479; CW-binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW-binding 1; 8.  
DR Pfam; PF03224; Glyco\_hydro\_70; 1.  
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 36  
FT CHAIN 37 1365  
FT REPEAT 157 177  
FT REPEAT 178 197  
FT DOMAIN 198 1061  
FT REPEAT 1062 1082  
FT REPEAT 1083 1102  
FT REPEAT 1150 1169  
FT REPEAT 1170 1190  
FT REPEAT 1225 1243  
FT REPEAT 1289 1308  
FT REPEAT 1309 1328  
FT REPEAT 1331 1352  
FT SEQUENCE 1365 AA; 151590 MW; 16729685A28C476 CRC64;

Query Match 62.2%; Score 69; DB 1; Length 1365;  
Best Local Similarity Pred. No. 0.01;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
1 VPSYSFIRAHDSVQDILIRMI 22  
||:|||||

Db 537 VPMYVIFIRAHDSVQTRIAKII 558

## RESULT 6

GTFD STRMU STANDARD; PRT; 1462 AA.  
AC P49331; 069383; 069386; 069389; 069392; 069398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucosyltransferase-5 precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN GTFD OR SMU. 910.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=GS-5;  
RX MEDLINE=91100958; PubMed=2148600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-5 enzyme.";  
RL U. Gen. Microbiol. 136:2099-2105 (1990).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,  
CC MT4467 / Serotype E, and MT8148 / Serotype C;  
CC MEDLINE=98231643; PubMed=9570124;  
CC Fujisawa T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
CC Kimura S., Hamada S.;  
CC "Molecular analyses of glucosyltransferase genes among strains of  
CC Streptococcus mutans.";  
CC FEWS Microbiol. Lett. 161:331-336 (1998).  
CC [3]  
CC SEQUENCE FROM N.A.  
CC STRAIN=UA159 / ATCC 700610 / Serotype C;  
CC MEDLINE=22295063; PubMed=12397186;  
CC Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Qian Y.,  
CC Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Perretti J.J.;  
CC Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Perretti J.J.;  
CC "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
CC pathogen.";  
CC Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl) (N) = D-  
fructose + ((1,6)-alpha-D-glucosyl) (N+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: Belongs to family 70 of glycosyl hydrolases.  
CC -1- SIMILARITY: Contains 6 cell wall binding repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M29296; AAA26895.1; -  
DR EMBL; D88653; BAA26103.1; -  
DR EMBL; D88656; BAA26107.1; -  
DR EMBL; D88659; BAA26111.1; -  
DR EMBL; D88662; BAA26115.1; -  
DR EMBL; D89979; BAA26121.1; -

DR EMBL; AE014932; AANS6619.1; -.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CM\_binding\_11.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries;  
 KW Complete proteome.  
 FT CHAIN 1 1462  
 FT SIGNAL 1 1462  
 FT DOMAIN 1232 1433  
 FT REPEAT 1232 1235  
 FT REPEAT 1236 1359  
 FT REPEAT 1360 1423  
 FT REPEAT 10 10  
 FT VARIANT 19 19  
 FT VARIANT 58 58  
 FT VARIANT 68 68  
 FT VARIANT 81 81  
 FT VARIANT 113 113  
 FT VARIANT 122 122  
 FT VARIANT 132 132  
 FT VARIANT 135 135  
 FT VARIANT 137 137  
 FT VARIANT 202 202  
 FT VARIANT 255 255  
 FT VARIANT 275 275  
 FT VARIANT 288 288  
 FT VARIANT 301 301  
 FT VARIANT 313 313  
 FT VARIANT 317 317  
 FT VARIANT 328 328  
 FT VARIANT 350 350  
 FT VARIANT 628 628  
 FT VARIANT 688 688  
 FT VARIANT 726 726  
 FT VARIANT 762 762  
 FT VARIANT 964 964  
 FT VARIANT 1019 1019  
 FT VARIANT 1059 1060  
 FT VARIANT 1060 1060  
 FT VARIANT 1080 1080  
 FT VARIANT 1142 1142  
 FT VARIANT 1198 1198  
 FT VARIANT 1220 1220  
 FT VARIANT 1280 1280  
 FT VARIANT 1282 1282  
 FT VARIANT 1290 1290  
 FT VARIANT 1311 1311  
 FT VARIANT 1403 1403  
 FT VARIANT 1425 1425  
 FT VARIANT 1449 1449  
 FT CONFLICT 1448 1462  
 FT SEQUENCE 1462 AA; 163387 MW; CE4A279CAD708645 CRC64;  
 Query Match 58.2%; Score 64; DB 1; Length 1462;  
 Best Local Similarity 65.0%; Pred. No. 0.063;  
 Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

RESULT 7  
 ID GALT\_BACHD STANDARD; PRT; 508 AA.  
 AC O9NDV2; O9RC74;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) (Gal-1-P  
 DE uridylyltransferase) (UDP-glucose--hexose-1-phosphate  
 DE uridylyltransferase).  
 GN GALT OR BH109.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R.,  
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuwara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 RN [2]  
 RP SEQUENCE OF 1-450 FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=99411980; PubMed=10484179;  
 RA Takami H., Takaki Y., Nakasone K., Sakiyama T., Maeno G., Sasaki R.,  
 RA Hirama C., Fujii F., Masui N.;  
 RT "Genetic analysis of the chromosome of alkaliphilic Bacillus  
 RT halodurans C-125.";  
 RL Extremophiles 3:227-233(1999).  
 CC -1- CARBOLYTIC ACTIVITY: UDP-glucose + alpha-D-galactose 1-phosphate =  
 CC alpha-D-glucose 1-phosphate + UDP-galactose.  
 CC -1- PATHWAY: Galactose metabolism; second step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: Belongs to the galactose-1-phosphate  
 CC uridylyltransferase family 2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB001510; BAB04828.1; -.  
 CC EMBL; AB024554; BAA83925.1; -.  
 CC PIR; E83788; E83788.  
 CC HAMAP; MF\_00571; -; 1.  
 DR InterPro; IPR000766; GalP\_transf\_II.  
 DR InterPro; IPR005850; GalP\_transf\_C.  
 DR InterPro; IPR005849; GalP\_transf\_N.  
 DR InterPro; IPR005934; GalT2.  
 DR Pfam; PF02744; GalP\_UDP\_tr\_C; 1.  
 DR Pfam; PF01087; GalP\_UDP\_transf; 1.  
 DR TIGRfam; TIGR01239; galT\_2; 1.  
 DR PROSITE; PS01163; GAL\_P\_UDP\_TRANSF\_II; 1.  
 KW Transferrase; Nucleotidyltransferase; Galactose metabolism;  
 KW Complete proteome.  
 FT SEQUENCE 508 AA; 57989 MW; 16A6F607FCEBA2E CRC64;  
 Query Match 44.5%; Score 49; DB 1; Length 508;  
 Best Local Similarity 62.5%; Pred. No. 4.9;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

RESULT 8
SYE_STABP STANDARD; FRT; 484 AA.
ID STABP STANDARD; FRT; 484 AA.
AC ORCTU3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glutamate-tRNA synthetase (EC 6.1.1.17) (Glutamate-tRNA ligase)
DE (Gurs).
GN GLUTX OR SE0290.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RC PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qian Z.-Q., Zhao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Men Y.-M.,
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamy]-tRNA(Glu).
CC -1- SUBUNIT: Monomer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminocyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016744; AAC03887.1; -
CC HAMAP; MF_00022; -; 1.
CC InterPro; IPR004527; GluX bact.
CC InterPro; IPR000924; Glu tRNA-synt_1c.
CC InterPro; IPR008925; tRNA-synt bind.
CC InterPro; IPR001412; tRNA-synt_1.
CC Pfam; PF00749; tRNA-synt_1c; 1.
CC PRINTS; PR00397; TRNASYNTHGLU.
CC TIGFAMs; TIGR00464; glxX_bact; 1.
CC PROSITE; PS00178; AA_tRNA_LIGASE_1; 1.
CC KMW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC FT SITE 11 21 "HIGH" REGION.
CC FT SITE 252 256 "KMSK" REGION.
CC FT BINDING 255 255 ATP (BY SIMILARITY).
CC SEQUENCE 484 AA; 56370 MW; 644A72F0CB824FEB CRC64;
OY
Query Match 42.3%; Score 46.5; DB 1; Length 484;
Best Local Similarity 45.5%; Pred. No. 11;
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;
DB 1 VPSYSPARA--HSEVQDIIR 19
193 VETYNFAVAVDDHYMQISDVIR 214

```

```

GN NPR3 OR ANPRC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Kidney;
RX MEDLINE=90287735; PubMed=2162522;
RA Lowe D.G., Camerato T.R., Goeddel D.V.;
RT "cDNA sequence of the human atrial natriuretic peptide clearance
RT receptor.";
RL Nucleic Acids Res. 18:3412-3412(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=9036656; PubMed=2169733;
RA Porter J.G., Arsten A., Fuller F., Miller J.A., Gregory L.C.,
RA Lewicki J.A.;
RT "Isolation and functional expression of the human atrial natriuretic
RT peptide clearance receptor cDNA.";
RL Biochem. Biophys. Res. Commun. 171:796-803(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Lens epithelium;
RA Rae J.L., Shepard A.R.;
RT "Human lens epithelial mRNA for atrial natriuretic peptide clearance
RT receptor.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE. DOES NOT HAVE
CC GUANYLATE CYCLASE ACTIVITY.
CC -1- SUBUNIT: Homodimer; disulfide-linked.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1; Isoform=2; Sequence=displayed;
CC Name=2; Isoform=1; Sequence=displayed;
CC -1- MISCELLANEOUS: THERE SEEM TO BE AT LEAST THREE AND RECEPTORS: TWO
CC WITH GUANYLATE CYCLASE ACTIVITY (ANP-A AND ANP-B) AND ONE (ANP-C)
CC WHICH IS PROBABLY RESPONSIBLE FOR THE CLEARANCE OF ANP FROM THE
CC CIRCULATION WITHOUT A ROLE IN SIGNAL TRANSDUCTION.
CC -1- SIMILARITY: TO ANP-A AND ANP-B RECEPTORS IN THEIR EXTRACELLULAR
CC AND TRANSMEMBRANE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52282; CA36523.1; -
CC EMBL; M59305; AA51734.1; -
CC EMBL; AF025998; AAB8801.1; -
CC PIR; S10150; OYHOCR.
CC PDB; 1JDN; 05-SEP-01.
CC Genew; HGNC:7945; NPR3.
CC MIM; 108962; -
CC GO; GO:0004872; F:receptor activity; TAS.
CC GO; GO:0001501; F:skeletal development; TAS.
CC InterPro; IPR001828; ANP_receptor.
CC InterPro; IPR001170; Ntpep_receptor.
CC Pfam; PF01094; ANP_receptor; 1.
CC PRINTS; PR00255; NATPEPTIDER.
CC PROSITE; PS00458; ANP_RECEPTORS; 1.
CC Receptor; Glycoprotein; Transmembrane; Signal; Alternative splicing;
CC 3D-structure.
CC FT SIGNAL 1 26 POTENTIAL.
CC FT CHAIN 27 541 ATRIAL NATRIURETIC PEPTIDE CLEARANCE
CC RECEPTOR.
CC FT DOMAIN 27 481 EXTRACELLULAR (POTENTIAL).

```





CX NCBI\_TaxID=158878, 158879, 196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ms50 / ATCC 700699, and N115;  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,  
 Kanamori M., Matsunari H., Maruyama A., Murakami H., Hoshiyama A.,  
 Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 Sekimizu K., Hirakawa H., Kishida S., Goto S., Yabuzaki J.,  
 Kanehisa M., Masahita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 Hatori M., Ogasawara N., Hayashi H., Hiratake K.;  
 RA "Whole genome sequencing of methicillin-resistant *Staphylococcus*  
 RT *aureus*.";  
 RL Lancet 357:1225-1240(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MM2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,  
 Yamamoto K., Hiratake K.;  
 RA "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827(2002).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
 CC diphosphate + L-glutamyl-tRNA(Glu).  
 CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-I aminacyl-tRNA synthetase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AP003359; BAB56690.1; -  
 CC EMBL; AP003130; BAB41716.1; -  
 CC EMBL; AP004823; BAB94348.1; -  
 CC PIR; A89820; A89820.  
 DR HSSP; P27000; 1GLN.  
 DR SWISS-PROT; P27000; STRAN.  
 DR HAMAP; MF 00022; -; 1.  
 DR InterPro; IPR004527; GLX bact.  
 DR InterPro; IPR000824; GLU tRNA-synt 1c.  
 DR InterPro; IPR008925; tRNA-synt 1c.  
 DR InterPro; IPR001412; tRNA-synt 1c.  
 DR Pfam; PF00749; tRNA-synt 1c; 1.  
 DR PRINTS; PR00987; TRNASYNTHGU.  
 DR TIGRFAMs; TIGR00464; GLX bact; 1.  
 DR PROSITE; PS00178; AA.TRNA\_LIGASE\_I; 1.  
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;  
 KW Complete proteome.  
 KM Complete proteome.  
 FT SITE 11 "HIGH" REGION.  
 FT SITE 252 "KMSKS" REGION.  
 FT BINDING 255 ATP (By similarity).  
 FT BINDING 255 ATP (By similarity).  
 SQ SEQUENCE 484 AA; 56288 MW; 4CB45F08DA238FA CRC64;  
 Query Match 41.4%; Score 45.5; DB 1; Length 484;  
 Best Local Similarity 40.9%; Pred. No. 16;  
 Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;  
 QY 1 VPSYSPARA---HDSVVDIIR 19  
 ID IDPTNFAVAIDDDHMQISDVIR 214  
 DB 193 IPTNFAVAIDDDHMQISDVIR 214  
 RESULT 12  
 SYE\_STR6 STANDARD; PRT; 486 AA.  
 ID IDPTNFAVAIDDDHMQISDVIR 214

AC ORCMN5;  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glutamy1-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)  
 DE (GluS).  
 GN GLTX OR SPRI881.  
 OS *Streptococcus pneumoniae* (strain ATCC BAA-255 / R6).  
 CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 CX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczyk L.C., Burgett S.,  
 Dehoff B.S., Bstrem S.T., Filtz L., Fu D.-O., Fuller W., Geringer C.,  
 RA Gilmour R., Glass J.S., Knofa H., Kraft A.R., Lagace R.E.,  
 RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
 RA McInnes S.M., McInnes M., McInnes K., Mundy C.W., Nicas T.I.,  
 RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,  
 RA Sun P.-W., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Usakinas S.R., Kostek P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RA "Genome of the bacterium *Streptococcus pneumoniae* strain R6.";  
 RT J. Bacteriol. 183:5709-5717(2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
 CC diphosphate + L-glutamyl-tRNA(Glu).  
 CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-I aminacyl-tRNA synthetase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE008552; AAL00683.1; -  
 CC PIR; F98106; F98106.  
 DR HAMAP; MF 00022; -; 1.  
 DR InterPro; IPR004527; GLX bact.  
 DR InterPro; IPR000924; GLU tRNA-synt 1c.  
 DR InterPro; IPR008925; tRNA-synt 1c.  
 DR InterPro; IPR001412; tRNA-synt 1c.  
 DR Pfam; PF00749; tRNA-synt 1c; 1.  
 DR PRINTS; PR00987; TRNASYNTHGU.  
 DR TIGRFAMs; TIGR00464; GLX bact; 1.  
 DR PROSITE; PS00178; AA.TRNA\_LIGASE\_I; 1.  
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;  
 KW Complete proteome.  
 KM Complete proteome.  
 FT SITE 11 "HIGH" REGION.  
 FT SITE 255 "KMSKS" REGION.  
 FT BINDING 258 ATP (By similarity).  
 FT BINDING 258 ATP (By similarity).  
 SQ SEQUENCE 466 AA; 55912 MW; F7315B21CB22381D CRC64;  
 Query Match 40.5%; Score 44.5; DB 1; Length 486;  
 Best Local Similarity 42.9%; Pred. No. 24;  
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;  
 QY 2 PYSYSPARA---HDSVVDIIR 19  
 ID IDPTNFAVAIDDDHMQISDVIR 216  
 DB 196 PTNFAVAIDDDHMQISDVIR 216  
 RESULT 13  
 UVRG LISIN STANDARD; PRT; 603 AA.  
 ID IDPTNFAVAIDDDHMQISDVIR 216  
 AC Q92CH5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)



DE UvrABC system protein C (UvrC protein) (Excinuclease ABC subunit C).  
 GN UVR\_C OR L1N1197.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 RX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaeser P., Frangeul L., Buchrieser C., Rusnok C., Amend A.,  
 BA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Chardot A., Chetoui F., Couve E., de Daruvar A., Deboux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,  
 RA Ertian K.-D., Fahl H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft U., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjati H.,  
 RA Nordleik G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT "Comparative genomics of *Listeria* species."; Science 294:849-852(2001).  
 CC -1- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 CC processing of DNA lesions. UvrC both incises the 5' and 3' sides  
 CC of the lesion. The N-terminal half is responsible for the 5'  
 CC incision and the C-terminal half is responsible for the 3'  
 CC incision (By similarity).  
 CC -1- SUBUNIT: Interacts with uvrB in an incision complex (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the uvrC family.  
 CC -1- SIMILARITY: Contains 1 Uvr domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL596167; CAC96428.1; -  
 DR PIR; AD1582; AD1582.  
 DR UniProt; L1N01197; -  
 DR HAMAP; MF\_00203; -; 1.  
 DR InterPro; IPR003583; HHH\_1.  
 DR InterPro; IPR001943; UvrB/C.  
 DR InterPro; IPR004791; UvrC.  
 DR InterPro; IPR001162; UvrC\_C.  
 DR Pfam; PF01541; Excl\_endo\_N\_1.  
 DR Pfam; PF02151; UVR\_1.  
 DR ProDom; PD005870; UvrC\_C; 1.  
 DR SMART; SM00465; G1YC; 1.  
 DR SMART; SM00278; HHH1; 1.  
 DR TIGRFAMs; TIGR00194; uvrC; 1.  
 DR PROSITE; PS50151; UVR; 1.  
 DR PROSITE; PS50164; UVR\_C\_1; 1.  
 DR PROSITE; PS50165; UVR\_C\_2; 1.  
 DR SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; Complete proteome.  
 KM DNA excision; Complete proteome.  
 FT DOMAIN 197 232 UVR.  
 SQ SEQUENCE 603 AA; 69302 MW; 8FA17669DB8EFA1 CRC64;

Query Match 40.5%; Score 44.5; DB 1; Length 603;  
 Best Local Similarity 40.9%; Pred. No. 30;  
 Matches 9; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

QY 1 VPVSFARAHSEVDDIIRDT 22  
 DB 42 VRSY-FSGTHDSKTQRLVGEIV 62

RESULT 14  
 ID UVR\_C LISMO STANDARD; PRT; 603 AA.  
 AC Q8V7P0;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE UvrABC system protein C (UvrC protein) (Excinuclease ABC subunit C).  
 GN UVR\_C OR L1N1234.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 RX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaeser P., Frangeul L., Buchrieser C., Rusnok C., Amend A.,  
 BA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Chardot A., Chetoui F., Couve E., de Daruvar A., Deboux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,  
 RA Ertian K.-D., Fahl H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft U., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjati H.,  
 RA Nordleik G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT "Comparative genomics of *Listeria* species."; Science 294:849-852(2001).  
 CC -1- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 CC processing of DNA lesions. UvrC both incises the 5' and 3' sides  
 CC of the lesion. The N-terminal half is responsible for the 5'  
 CC incision and the C-terminal half is responsible for the 3'  
 CC incision (By similarity).  
 CC -1- SUBUNIT: Interacts with uvrB in an incision complex (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the uvrC family.  
 CC -1- SIMILARITY: Contains 1 Uvr domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL591978; CAC99312.1; -  
 DR PIR; AB1229; AB1229.  
 DR UniProt; L1M001234; -  
 DR HAMAP; MF\_00203; -; 1.  
 DR InterPro; IPR003583; HHH\_1.  
 DR InterPro; IPR001943; UvrB/C.  
 DR InterPro; IPR004791; UvrC.  
 DR InterPro; IPR001162; UvrC\_C.  
 DR InterPro; IPR00305; UvrC\_N.  
 DR Pfam; PF01541; Excl\_endo\_N\_1.  
 DR Pfam; PF02151; UVR\_1.  
 DR ProDom; PD005870; UvrC\_C; 1.  
 DR SMART; SM00465; G1YC; 1.  
 DR SMART; SM00278; HHH1; 1.  
 DR TIGRFAMs; TIGR00194; uvrC; 1.  
 DR PROSITE; PS50151; UVR; 1.  
 DR PROSITE; PS50164; UVR\_C\_1; 1.  
 DR PROSITE; PS50165; UVR\_C\_2; 1.  
 DR SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; Complete proteome.  
 KM DNA excision; Complete proteome.  
 FT DOMAIN 197 232 UVR.  
 SQ SEQUENCE 603 AA; 69315 MW; E87D742AF5F325B CRC64;

Query Match 40.5%; Score 44.5; DB 1; Length 603;  
 Best Local Similarity 40.9%; Pred. No. 30;

Matches 9; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

QY 1 VPSYFARADSEVQDIIRDI 22  
 DB 42 VRSY-FSGTDSKTRVQYIV 62

RESULT 15  
 YH75 ARCFU STANDARD; PRT; 330 AA.  
 AC 028499;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Hypothetical protein AF1775.  
 GN AF1775.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Kleink H.-P., Clayson R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kervagane A.R., Graham D.E., Kyprides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Meldrum J.F., McDonald L., Uterback T.,  
 RA Cotton M.D., Spillig T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujil C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -1- SIMILARITY: Belongs to the ATZ/TRZ family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE000980; AAB89475.1; -;  
 DR PIR; F69471; F69471.  
 DR TIGR; AF1775; -;  
 DR InterPro: IPR0006680; Amidohydro\_1.  
 DR Pfam; PF01979; Amidohydro\_1; 1.  
 DR Hypothetical protein; Hydrolase; Complete proteome.  
 KW SEQUENCE 330 AA; 37069 MW; 525AD9F7F35A6FB1 CRC64;  
 SQ

Query Match 40.0%; Score 44; DB 1; Length 330;  
 Best Local Similarity 42.1%; Pred. No. 19;  
 Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARADSEVQDIIRDI 21  
 DB 152 AYSSARDHDKLMEVREI 170

RESULT 16  
 RROC WEIHE STANDARD; PRT; 1046 AA.  
 AC P96177;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase

DE beta' chain) (RNA polymerase beta' subunit) (Fragment).  
 GN RROC.  
 OS Weissella hellenica.  
 OC Bacteria; Firmicutes; Lactobacillales; Weissella.  
 OX NCBI\_TaxID=46256;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCFB 2973;  
 RX MEDLINE=97016803; PubMed=8863429;  
 RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.,  
 RT "Analysis of the beta' subunit of DNA-dependent RNA polymerase does  
 RT not support the hypothesis inferred from 16S rRNA analysis that  
 RT Genococcus oeni (formerly Leuconostoc oenos) is a tachyelic  
 RT (fast-evolving) bacterium.";  
 RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).  
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The enzyme consists of the sigma chain and the core  
 CC enzyme which is composed of 2 alpha chains, 1 beta chain, and 1  
 CC beta' chain.  
 CC -1- SIMILARITY: Belongs to the RNA polymerase beta' chain family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X96470; CA65322.1; -;  
 DR HSSP; Q9XWU6; IHQM.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR007080; RNA\_pol\_Rpb1\_1.  
 DR InterPro: IPR007066; RNA\_pol\_Rpb1\_3.  
 DR InterPro: IPR007083; RNA\_pol\_Rpb1\_4.  
 DR InterPro: IPR007081; RNA\_pol\_Rpb1\_5.  
 DR InterPro: IPR006592; RNA\_pol\_A\_N.  
 DR Pfam; PF04997; RNA\_pol\_Rpb1\_1; 1.  
 DR Pfam; PF00623; RNA\_pol\_Rpb1\_2; 1.  
 DR Pfam; PF04983; RNA\_pol\_Rpb1\_3; 1.  
 DR Pfam; PF05000; RNA\_pol\_Rpb1\_4; 1.  
 DR Pfam; PF04998; RNA\_pol\_Rpb1\_5; 1.  
 DR SMART; SM00663; RPOA\_N; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription.  
 FT NON\_TER 1 1  
 FT NON\_TER 1046 1046  
 FT SEQUENCE 1046 AA; 117107 MW; D54C62C26A7F1696 CRC64;  
 SQ

Query Match 40.0%; Score 44; DB 1; Length 1046;  
 Best Local Similarity 46.2%; Pred. No. 65;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 10 HDSEVQDIIRDI 22  
 DB 327 HDEVDVDTLBDVI 339

RESULT 17  
 POL3\_DROME STANDARD; PRT; 1058 AA.  
 AC P04323;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Retrovirus-related Pol polyprotein from transposon 17.6 (Contains:  
 DE Protease (EC 3.4.23.-); Reverse transcriptase (EC 2.7.7.49);  
 DE Endonuclease).  
 GN POL.  
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RX NCBI\_TaxID=7227;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85061628; PubMed=6209583;  
 RA Saito K., Kugimiyama W., Matsuo Y., Inouye S., Yoshioka K., Yuki S.;  
 RT "Identification of the coding sequence for a reverse  
 transcriptase-like enzyme in a transposable genetic element in  
 RT *Drosophila melanogaster*.";  
 RL Nature 312:659-661(1984).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + [DNA] (N).  
 CC -1- MISCELLANEOUS: The open reading frame is located in a copia-like  
 CC transposable element called 17.6.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, X01472; CAA25702.1; -.  
 CC PIR, A03971; GNF17.  
 DR MEROPS: A02.052; -.  
 DR FlyBase, FBgn0014453; 17.6/pol.  
 DR InterPro: IPR001968; Asparticase\_AS.  
 DR InterPro: IPR009007; Pept\_Acid.  
 DR InterPro: IPR001995; Peptidase\_A2.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR000477; RVTse.  
 DR Pfam, PF00665; Rve; 1.  
 DR Pfam, PF00077; RVP; 1.  
 DR Pfam, PF00078; Rvt; 1.  
 DR PROSITE, PS00141; ASP\_PROTEASE; 1.  
 KM Hydroxylase; Aspartyl protease; RNA-directed DNA polymerase;  
 KM Hydrolase; Aspartyl protease; Polypeptide; Transposable element.  
 FT ACT\_SITE 30 PROTEASE (BY SIMILARITY).  
 SQ SEQUENCE 1058 AA; 122697 MW; C833F5C4A7E1F091 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 1058;  
 Best Local Similarity 36.8%; Pred. No. 66;  
 Matches 7; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 4 YSFARAHDSVQDITRDI 22  
 DB 213 YSYQAYEVEESQIDMT 231

RESULT 18  
 ID UBIE RHILLO STANDARD; PRT; 258 AA.  
 AC Q98GV1;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ubiquitinone/menquinone biosynthesis methyltransferase ubie  
 DE (EC 2.1.1.-).  
 GN UBIE OR MUR3165.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Mesorhizobium.  
 OC NCBI\_TaxID=381;  
 RX SEQUENCE FROM N.A.  
 RP STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shingo S., Sugimoto M.,  
 RA Takuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT *Mesorhizobium loti*.";  
 RL DNA Res. 7:331-338(2000).  
 CC -1- FUNCTION: Methyltransferase required for the conversion of  
 CC dimethylmenaquinone (DMKH2) to menaquinone (MKH2) and the  
 CC conversion of 2-polyphenyl-6-methoxy-1,4-benzoquinol (DDMQH2) to  
 CC 2-polyphenyl-3-methyl-6-methoxy-1,4-benzoquinol (DMQH2) (By  
 CC similarity).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 2-polyphenyl-6-  
 CC methoxy-1,4-benzoquinol = S-adenosyl-L-homocysteine + 2-  
 CC polyphenyl-3-methyl-6-methoxy-1,4-benzoquinol.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + demethylmenaquinol =  
 CC S-adenosyl-L-homocysteine + menaquinol.  
 CC -1- PATHWAY: Menaquinone biosynthesis; last step.  
 CC -1- PATHWAY: Ubiquinone biosynthesis.  
 CC -1- SIMILARITY: Belongs to the ubiB family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, AF003001; BAB50115.1; -.  
 CC HAMAP, MF\_01813; -; 1.  
 DR InterPro: IPR001601; Methyltransf.  
 DR InterPro: IPR000051; SAM bind.  
 DR InterPro: IPR004034; Ubi/Men Methyltransf.  
 DR InterPro: IPR004033; UbiS/COQ5\_Methyltransf.  
 DR Pfam, PF01209; UbiL\_methyltran; 1.  
 DR PROSITE, PS01183; UBIL\_1; 1.  
 DR PROSITE, PS01184; UBIL\_2; 1.  
 KM Menaquinone biosynthesis; Ubiquinone biosynthesis; Transferase;  
 KM Methyltransferase; Complete proteome.  
 SQ SEQUENCE 258 AA; 28291 MW; 9909682B725B2828 CRC64;

Query Match 39.1%; Score 43; DB 1; Length 258;  
 Best Local Similarity 36.8%; Pred. No. 20;  
 Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 YSFARAHDSVQDITRDI 21  
 DB 15 SYGFKRVRGEGKQSLVNDV 33

RESULT 19  
 ID Y326 MYCGB STANDARD; PRT; 295 AA.  
 AC P47568;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical UPF0230 protein MG326.  
 GN MG326.  
 OS Mycoplasma genitalium.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmales; Mycoplasma.  
 OC NCBI\_TaxID=2097;  
 RX SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE=96026346; PubMed=7569993;  
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandhu M., Fuhmann J.L.,  
 RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
 RA Tomb J.-F., Dougherty B.A., Hutchison C.A. III, Venter J.C.,  
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 RT "The minimal gene complement of *Mycoplasma genitalium*.";

RL Science 270:397-403(1995).  
CC -1- SIMILARITY: BELONGS TO THE UPF0230 (DEGV) FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U39714; AAC71550.1; -.  
DR PIR: A64236; A64236.  
DR TIGR: MG326; -.  
DR InterPro: IPR003797; DegV.  
DR Pfam: PF02645; DUF194; 1.  
DR TIGRfam: TIGR00762; DegV, 1.  
DR Hypothetical protein; Complete proteome.  
RW SEQUENCE 295 AA; 33413 MW; 04610881C0F841BE CRC64;  
SQ  
  
Query Match 39.1%; Score 43; DB 1; Length 295;  
Best Local Similarity 42.1%; Pred. No. 23;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
  
4 YSPFARHDSVQDITRDI 22  
||| : : : : :  
Db 239 YSPCKNYANEIKITIDPI 257  
  
RESULT 20  
MITE MOUSE STANDARD: PRT; 526 AA.  
ID M08874; 008885; 008803; 008843; 060782; 060782; 09J10; 09J11;  
AC 09J12; 09J13; 09J14; 09J15; 09J16; 09JX9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Microphthalmia-associated transcription factor.  
GN MTF OR Mf OR BM OR VIT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS M AND M1), AND VARIANTS M1 AND M1-WS.  
RC TISSUE=Melanocyte;  
RX MEDLINE=93345026; PubMed=8343963;  
RA Hodgkinson C.A., Moore K.J., Nakayama A., Steingrimsen E.,  
RA Copeland N.G., Jenkins N.A., Arnheiter H.,  
RT "Mutations at the mouse microphthalmia locus are associated with  
RT defects in a gene encoding a novel basic-helix-loop-helix-zipper  
RT protein."  
RL Cell 74:395-404(1993).  
[2]  
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS.  
RC STRAIN=129/Sv; TISSUE=Heart;  
RX MEDLINE=20253112; PubMed=10790403;  
RA Hallison J.H., Favor J., Hodgkinson C., Glaser T., Lamoreux M.L.,  
RA Jenkins N.A., Steingrimsen E.,  
RT "Genomic, transcriptional and mutational analysis of the mouse  
RT microphthalmia locus."  
RL Genetics 155:291-300(2000).  
[3]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS H AND M), AND VARIANTS.  
RC STRAIN=C57BL/6; TISSUE=Heart, and Melanocyte;  
RX MEDLINE=95179171; PubMed=7874168;  
RA Steingrimsen E., Moore K.J., Lamoreux M.L., Ferre-D'Amaré A.R.,  
RA Burley S.K., Sanders Zimring D.C., Skow L.C., Hodgkinson C.A.,  
RA Arnheiter H., Copeland N.G., Jenkins N.A.,  
RT "Molecular basis of mouse microphthalmia (mi) mutations helps explain  
RT their developmental and phenotypic consequences."  
RL Nat. Genet. 8:256-263(1994).

RN [4]  
RP SEQUENCE OF 345-392 FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Heart;  
RX MEDLINE=94012591; PubMed=8407885;  
RA Hughes M.J., Lingrel J.B., Krakowsky J.M., Anderson K.P.;  
RT "A helix-loop-helix transcription factor-like gene is located at the  
RT m1 locus."  
RL J. Biol. Chem. 268:20687-20690(1993).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=98321192; PubMed=9647758;  
RA Aase S., Fuxe N., Yasumoto K.-I., Sato S., Yajima I., Yamamoto H.,  
RA Udono T., Durlu Y.K., Tamai M., Takahashi K., Shibahara S.;  
RT "Identification of a novel isoform of microphthalmia-associated  
RT transcription factor that is enriched in retinal pigment epithelium."  
RL Biochem. Biophys. Res. Commun. 247:710-715(1998).  
RN [6]  
RP VARIANT M1-BW.  
RX MEDLINE=99310550; PubMed=10400990;  
RA Yajima I., Sato S., Kimura T., Yasumoto K.-I., Shibahara S.,  
RA Goding C.R., Yamamoto H.;  
RT "An 11 element intronic insertion in the black-eyed white (Mltfmi-bw)  
RT gene: the loss of a single Mltf isoform responsible for the  
RT pigmentary defect and inner ear deafness."  
RL Hum. Mol. Genet. 8:1431-1441(1999).  
RN [7]  
RP SUBCELLULAR LOCATION.  
RX MEDLINE=96182124; PubMed=8622664;  
RA Takebayashi K., Chida K., Tsukamoto I., Morii E., Munakata H.,  
RA Arnheiter H., Kuroki T., Kitamura Y., Nomura S.;  
RT "The recessive phenotype displayed by a dominant negative  
RT microphthalmia-associated transcription factor mutant is a result of  
RT impaired nucleation potential."  
RL Mol. Cell. Biol. 16:1203-1211(1996).  
CC -1- FUNCTION: Transcription factor for tyrosinase and tyrosinase-  
CC related protein 1. Binds to a symmetrical DNA sequence (E-Boxes)  
CC (5'-CACGGG-3') found in the tyrosinase promoter. Plays a critical  
CC role in the differentiation of various cell types as neural crest-  
CC derived melanocytes, mast cells, osteoclasts and optic cup-derived  
CC retinal pigment epithelium.  
CC -1- SUBUNIT: Efficient DNA binding requires dimerization with another  
CC bHLH protein. Binds DNA in the form of homodimer or heterodimer  
CC with either TFEB, TFEB or TFEC.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=9;  
CC Name=A;  
CC IsoId=Q08874-1; Sequence=Displayed;  
CC Name=A1;  
CC IsoId=Q08874-2; Sequence=VSP\_002133;  
CC Name=A2;  
CC IsoId=Q08874-3; Sequence=VSP\_002131, VSP\_002134, VSP\_002136;  
CC Name=H;  
CC IsoId=Q08874-4; Sequence=VSP\_002129;  
CC Name=H1;  
CC IsoId=Q08874-5; Sequence=VSP\_002129, VSP\_002132;  
CC Name=H2;  
CC IsoId=Q08874-6; Sequence=VSP\_002129, VSP\_002132, VSP\_002135;  
CC Name=H3;  
CC IsoId=Q08874-7; Sequence=VSP\_002129, VSP\_002133;  
CC Name=M;  
CC IsoId=Q08874-8; Sequence=VSP\_002130;  
CC Name=M1;  
CC IsoId=Q08874-9; Sequence=VSP\_002130, VSP\_002134;  
CC -1- TISSUE SPECIFICITY: In the adult, expressed at high levels in the  
CC heart, skin, skeletal muscle, intestine, stomach, kidney, ovary,  
CC lung, spleen and brain. In the embryo, expressed in developing  
CC eye, ear, skin and heart. Isoform M1 is expressed in melanocytes  
CC and also in the embryonic and adult heart while isoforms A and H  
CC are more widely expressed.  
CC -1- PTM: PHOSPHORYLATION AT SER-405 SIGNIFICANTLY ENHANCES THE ABILITY  
CC TO BIND THE TYROSINASE PROMOTER (BY SIMILARITY).  
CC -1- DISEASE: DEFECTS IN MTF ARE THE CAUSE OF MICROPTHALMIA (MT), A

Query Match	Similarity	Score	DB	Length
Best Local	42.1%	43	1	526
Matches	8: Conservative	4: Mismatches	7: Indels	0: Gaps

2 PSYSEFAHDESVODIIR 20

469 PAVSIFKMGSNMEDILMD 487

RESULT 21

FUMA\_ECOLI STANDARD; PRT; 547 AA.

ID FUMA\_ECOLI STANDARD; PRT; 547 AA.

AC P00923; P76889; (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 31, Last sequence update)

DT 01-FEB-1995 (Rel. 41, Last annotation update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE Fumarate hydratase class I, aerobic (EC 4.2.1.2) (Fumarase).

GN FUMA OR B1612 OR C2004.

OS Escherichia coli. and

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

NCBI\_TaxID=562, 217992;

[1]

SEQUENCE FROM N.A.

MDLINE=84221385; PubMed=6328431,

Miles J.S., Guest J.R.,

"Complete nucleotide sequence of the fumarase gene fuma, of

Escherichia coli.";

Nucleic Acids Res. 12:3631-3642(1984).

[2]

SEQUENCE FROM N.A.

STRAIN=K12 / MG1655;

MDLINE=97426617; PubMed=9278503;

Blattner P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

Mau B., Shao Y.,

"The complete genome sequence of Escherichia coli K-12.";

Science 277:1453-1474(1997).

[3]

SEQUENCE FROM N.A.

STRAIN=K12;

MDLINE=97551357; PubMed=9097039;

Alba H., Baba T., Fujita K., Hayaishi K., Inada T., Isono K.,

Itch T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,

Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,

Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,

Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,

Tagamai H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,

Yamamoto Y., Horiuchi T.,

"A 570-bp DNA sequence of the Escherichia coli K-12 genome

corresponding to the 28.0-40.1 min region on the linkage map.";

DNA Res. 3:363-377(1996).

[4]

SEQUENCE FROM N.A.

STRAIN=O6:HI / CFT073 / ATCC 700928;

MDLINE=22388234; PubMed=12471157;

Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Dombrowsky M.S., Blattner F.R.,  
 RA "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 RM [5]  
 RP SEQUENCE OF 1-11.  
 RC STRAIN=K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.J., Robison K., Church G.M.,  
 RT "Comparing the predicted and observed properties of proteins encoded  
 RL in the genome of *Escherichia coli* K-12.";  
 RM Electrochim. Acta 18:1259-1313(1997).  
 RP [6]  
 RP IDENTIFICATION OF THE STRUCTURAL GENE.  
 RX MEDLINE=86142617; PubMed=3005475;  
 RA Guest J.R., Miles J.S., Roberts R.E., Woods S.A.,  
 RT "The fumarate genes of *Escherichia coli*: location of the *fumB* gene  
 RL and discovery of a new gene (*fumC*)."  
 RM J. Gen. Microbiol. 131:2971-2984(1985).  
 RP [7]  
 RP BIOCHEMICAL ANALYSIS OF FUMARATE AND FUMATE.  
 RX MEDLINE=88193096; PubMed=3282546;  
 RA Woods S.A., Schwartz S.D., Guest J.R.,  
 RT "Two biochemically distinct classes of fumarate in *Escherichia coli*.";  
 RL Biochim. Biophys. Acta 954:14-26(1988).  
 RP [8]  
 RP IRON-SULFUR CLUSTER.  
 RA Flint D.H., Emptage M.H., Guest J.R.,  
 RT "Fumarate A from *E. coli* contains a [4Fe-4S] cluster.";  
 RL J. Inorg. Biochem. 36:306-306(1989).  
 CC -1- FUNCTION: IT ACCOUNTS FOR ABOUT 80% OF THE FUMARATE ACTIVITY WHEN  
 CC CYCLE. IT ACCOUNTS FOR ABOUT 80% OF THE FUMARATE ACTIVITY WHEN  
 CC THE BACTERIA GROWS AEROBICALLY.  
 CC -1- CATALYTIC ACTIVITY: (S)-malate + fumarate + H(2)O.  
 CC -1- COFACTOR: Binds 1 4Fe-4S cluster.  
 CC -1- ENZYME REGULATION: SUBJECT TO AEROBIC RESPIRATORY CONTROL AND  
 CC CATABOLITE REPRESSION.  
 CC -1- PATHWAY: Tricarboxylic acid cycle.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SIMILARITY: TO OTHER THERMOLABILE CLASS I FUMARASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, X00532; CAA25204.1; -;  
 DR EMBL, AE000256; AAC74684.1; -;  
 DR EMBL, D90805; BAA15364.1; -;  
 DR EMBL, D90804; BAA15360.1; -;  
 DR EMBL, D90803; BAA15350.1; -;  
 DR EMBL, AE016761; AAN80464.1; -;  
 DR PIR, A03531; UFECAO.  
 CC Ecogene: EG10356; *fumA*.  
 DR InterPro: IPR000362; Fumarate\_lyase.  
 DR InterPro: IPR004646; TcdA\_fumA\_fumB.  
 DR InterPro: IPR004647; TcdB\_fumA\_fumB.  
 DR Pfam: PF05681; Fumarate\_C; 1.  
 DR Pfam: PF05683; Fumarate\_C; 1.  
 DR TIGRfam: TIGR00722; tcdA\_fumA\_fumB; 1.  
 DR TIGRfam: TIGR00723; tcdB\_fumA\_fumB; 1.  
 DR PROSITE: PS00163; FUMARATE\_LYASES; 1.  
 DR Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S;  
 KM Complete proteome.  
 FT INIT MET 0  
 FT METAL 317  
 FT ACT SITE 396  
 FT BINDING 462  
 FT SEQUENCE 547 AA; 60167 MW; F9827451050334D8 CRC64;  
 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 POTENTIAL.  
 SUBSTRATE CARBOXYL (POTENTIAL).  
 F9827451050334D8 CRC64;

Query Match 39.1%; Score 43; DB 1; Length 547;  
 Best Local Similarity 44.4%; Pred. No. 46;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
 QY 3 SYSPARADSEVDIIRD 20  
 DB 58 SFMRPAAQGVADILRD 75  
 RESULT 22  
 ID WR20\_ARATH STANDARD; PRT; 557 AA.  
 AC Q93W0; Q93W0; Q93W0; Q93W0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Probable WRKY transcription factor 20 (WRKY DNA-binding protein 20).  
 GN WRKY20 OR AT4G26640 OR T1SN24.90.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambut R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgatter M., de Simone V., Obermayer B., Maché R., Mueller M.,  
 RA Krels M., Delany M., Fulgomech P., Watson M., Schmidheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Bouty M., Bancroft I.,  
 RA Vos P., Hohnsbeil J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymoprez B., Chang Y.-D., Vandenbussche F.,  
 RA Braeken M., Welfens I., Voet M., Bastiens I., Aert R., Defoor E.,  
 RA Welzenegger T., Botche G., Ransperger U., Hilbert H., Braun M.,  
 RA Holzner B., Brandt A., Peters S., van Staveren M., Dirse W.,  
 RA Moijman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,  
 RA Bernier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,  
 RA Clark U., Doggett U., Hall S., Kay M., Lemard N., McIlroy K., Mayes R.,  
 RA Petrett A., Rajandream M.A., Lyne M., Benes V., Reumann S.,  
 RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Grandjean K., Pitaevski E.,  
 RA Neumann S., Argitau F., Clabaud G., Mendenhall A., Reiser R.,  
 RA Massenet O., Quigley F., Schmidt W., Lecharny A., Aubourg S.,  
 RA Schnabl S., Hiller R., Berger C., Montfort A., Casacuberta E.,  
 RA Chetoui F., Cooke R., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Furnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,  
 RA Frishman D., Haase D., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schütz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh U.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton L., Miller N., Greco T., Kemp K.,  
 RA Kramon U., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,  
 RA Nelson U., Speith J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,













OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxId=9913;  
 RN [1]  
 RX SEQUENCE FROM N.A. (ISOFORM 1).  
 RA MEDLINE=88243827; PubMed=2837487;  
 RA Fuller F., Porter J.G., Arsten A.E., James J.M., Schilling J.W.,  
 RA Scarborough R.M., Lewicki J.A., Schenk D.B.;  
 RT "Atrial natriuretic peptide clearance receptor. Complete sequence and  
 RT functional expression of cDNA clones";  
 RL J. Biol. Chem. 263:9395-9401(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=91250421; PubMed=1645726;  
 RA Saberi T., Mizuno T., Iwata T., Saito Y., Nagasawa T.,  
 RA Mizuno K.U., Ito F., Ito T., Hagiwara H., Hirose S.;  
 RT "Structure of the bovine atrial natriuretic peptide receptor (type C)  
 RT gene";  
 RL J. Biol. Chem. 266:11122-11125(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Lung;  
 RA MEDLINE=93186831; PubMed=8444892;  
 RA Mizuno T., Iwashina M., Itakura M., Hagiwara H., Hirose S.;  
 RT "A variant form of the type C atrial natriuretic peptide receptor  
 RT generated by alternative RNA splicing";  
 RL J. Biol. Chem. 268:5162-5167(1993).  
 RN [4]  
 RP SEQUENCE OF 151-179; 310-325 AND 446-452.  
 RX TISSUE=Lung;  
 RA MEDLINE=90088409; PubMed=2557006;  
 RA Uchida K., Mizuno T., Shimomaka M., Sugiyama N., Nara K., Ling N.,  
 RA Hagiwara H., Hirose S.;  
 RT "Purification and properties of active atrial-natriuretic-peptide  
 RT receptor (type C) from bovine lung";  
 RL Biochem. J. 263:671-678(1989).  
 RN [5]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=94179211; PubMed=8132555;  
 RA Itakura M., Iwashina M., Mizuno T., Ito T., Hagiwara H., Hirose S.;  
 RT "Molecular analysis of disulfide bridges in the type C atrial  
 RT natriuretic peptide receptor";  
 RL J. Biol. Chem. 269:8314-8318(1994).  
 CC -1- FUNCTION: RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE. DOES NOT HAVE  
 CC GUANYLATE CYCLASE ACTIVITY.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=PI0730-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=PI0730-2; Sequence=VSP 001811;  
 CC -1- MISCELLANEOUS: THERE SEEM TO BE AT LEAST THREE ANP RECEPTORS: TWO  
 CC WITH GUANYLATE CYCLASE ACTIVITY (ANP-A AND ANP-B) AND ONE (ANP-C)  
 CC WHICH IS PROBABLY RESPONSIBLE FOR THE CLEARANCE OF ANP FROM THE  
 CC CIRCULATION WITHOUT A ROLE IN SIGNAL TRANSDUCTION.  
 CC -1- SIMILARITY: TO ANP-A AND ANP-B RECEPTOR IN THEIR EXTRACELLULAR AND  
 CC TRANSMEMBRANE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J03876; AAA0376.1; -;  
 DR EMBL; D90372; BAA14380.1; -;  
 DR EMBL; D90365; BAA14380.1; JOINED.  
 DR EMBL; D90366; BAA14380.1; JOINED.  
 DR EMBL; D90367; BAA14380.1; JOINED.  
 DR EMBL; D90368; BAA14380.1; JOINED.

DR EMBL; D90369; BAA14380.1; JOINED.  
 DR EMBL; D90370; BAA14380.1; JOINED.  
 DR EMBL; D90371; BAA14380.1; JOINED.  
 DR EMBL; D13508; BAA02726.1; -;  
 DR PIR; A28101; A28111.  
 DR PIR; A45409; A45409.  
 DR InterPro; IPR001828; ANP\_receptor.  
 DR InterPro; IPR001170; Ntpep\_receptor.  
 DR Pfam; PF01094; ANP\_receptor; 1.  
 DR PRINTS; PRO0255; NATPEPTIDER.  
 DR PROSITE; PS00458; ANP\_RECEPTORS; 1.  
 DR Receptor; Glycoprotein; Transmembrane; Signal; Alternative splicing.  
 KM SIGNAL 1 20  
 FT PROPEP 21 41  
 FT CHAIN 42 537  
 FT DOMAIN 42 477  
 FT TRANSMEM 478 500  
 FT DOMAIN 501 537  
 FT DISULFID 104 132  
 FT DISULFID 209 257  
 FT DISULFID 469 469  
 FT CARBOHYD 82 82  
 FT CARBOHYD 289 289  
 FT CARBOHYD 465 465  
 FT VARSPLIC 472 473  
 FT CONFLICT 324 324  
 FT FT 537 59765 MW; B6AC8950FE8E996 CRC64;  
 SQ SREGION 537 AA; 59765 MW; B6AC8950FE8E996 CRC64;  
 Query Match 38.2% Score 42; DB 1; Length 537;  
 Best Local Similarity 36.8%; Pred. No. 65;  
 Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;  
 QY 3 SYSFARADSEVODIRDI 21  
 Db 229 AYNFDETKDIDLDIVRHI 247

Search completed: May 4, 2004, 09:10:42  
 Job time : 9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:05:56 ; Search time 33 Seconds  
(without alignments)  
210.345 Million cell updates/sec

Title: US-09-290-049A-19

Perfect score: 110  
Sequence: 1 VPSYFARADSEVQDIRDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

SPTREMBL 25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_plant:\*  
10: sp\_protist:\*  
11: sp\_rhodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1590	2	059983 streptococc
2	110	100.0	1590	2	055263 streptococc
3	85	77.3	1290	2	048756 leuconostoc
4	82	74.5	1330	2	084CN4 leuconostoc
5	82	74.5	1477	2	091466 leuconostoc
6	82	74.5	1508	2	09E2H5 leuconostoc
7	82	74.5	1508	2	052224 leuconostoc
8	81	73.6	1016	2	09ICJ7 leuconostoc
9	81	73.6	1527	2	09ZARA leuconostoc
10	81	73.6	1527	2	08RRE1 leuconostoc
11	80	72.7	1554	2	08KZL5 streptococc
12	75	69.1	1512	2	09WXS5 streptococc
13	75	68.2	2835	2	08G902 leuconostoc
14	72	65.5	1575	2	09ICJ3 streptococc
15	72	65.5	1577	2	054178 streptococc
16	72	65.5	1599	2	000599 streptococc

17	70	63.6	1449	2	068542 streptococc
18	70	62.7	1449	2	055264 streptococc
19	69	62.7	1338	2	09WJ4 streptococc
20	68	61.8	2057	2	09R05 leuconostoc
21	67	60.9	1518	2	000600 streptococc
22	64	58.2	1577	2	055265 streptococc
23	51	46.4	108	16	P74028
24	50	45.5	93	2	0921X9
25	50	45.5	1554	3	08J026
26	50	45.5	1567	3	08J0W2
27	47	42.7	303	16	08A338
28	47	42.7	449	16	08R03
29	47	42.7	533	2	08RUV0
30	47	42.7	881	16	08G5W2
31	46.5	42.3	470	12	08U2G0
32	46	41.8	200	16	08XJY7
33	46	41.8	367	4	09H769
34	46	41.8	506	4	09NS15
35	46	41.8	566	12	08OPY3
36	46	41.8	743	4	09NVJ7
37	46	41.8	743	4	09NV74
38	46	41.8	743	4	09BUN0
39	46	41.8	779	4	09H3P4
40	46	41.8	2219	4	08C0A3
41	46	41.8	2296	4	09Y3S1
42	45	40.9	51	16	09RNP7
43	45	40.9	147	10	07XIA0
44	45	40.9	209	16	082TR1
45	45	40.9	255	4	096BD8
46	45	40.9	462	16	0983M9
47	45	40.9	501	5	0964R1
48	45	40.9	597	16	09PR58
49	45	40.9	631	16	098PL8
50	45	40.9	1480	10	07XLJ7
51	45	40.9	1616	10	07XW87
52	44.5	40.5	603	4	08N608
53	44.5	40.5	2470	5	07YI02
54	44	40.0	179	12	09Q8W3
55	44	40.0	188	12	09Q815
56	44	40.0	188	12	083655
57	44	40.0	216	16	08Y5G9
58	44	40.0	257	17	08TYM8
59	44	40.0	316	10	08YTB0
60	44	40.0	518	10	09FTE2
61	44	40.0	527	10	08LGM8
62	44	40.0	1532	10	07XWU3
63	43	39.1	68	16	08YAI0
64	43	39.1	176	16	09XD61
65	43	39.1	282	16	07VU54
66	43	39.1	377	11	070241
67	43	39.1	385	17	08T262
68	43	39.1	467	16	08PL37
69	43	39.1	467	16	08P9D0
70	43	39.1	501	16	08UR52
71	43	39.1	548	16	08XAP8
72	43	39.1	548	16	08Z6R5
73	43	39.1	548	16	083ML7
74	43	39.1	2627	4	099973
75	42.5	38.6	104	2	09X771
76	42.5	38.6	688	10	041461
77	42	38.2	179	16	083GPD
78	42	38.2	217	8	037435
79	42	38.2	221	5	081937
80	42	38.2	269	16	081058
81	42	38.2	327	2	09X303
82	42	38.2	353	5	09V164
83	42	38.2	358	10	004546
84	42	38.2	361	5	08R24
85	42	38.2	368	16	0882W4
86	42	38.2	441	16	088A12
87	42	38.2	469	17	08TR12
88	42	38.2	480	5	095W16
89	42	38.2	480	13	090220

068542 streptococc	095264 streptococc	09WJ4 streptococc	09R05 leuconostoc	000600 streptococc	055265 streptococc	P74028	0921X9	08J026	08J0W2	08A338	08R03	08RUV0	08G5W2	08U2G0	08XJY7	09H769	09NS15	08OPY3	09NVJ7	09NV74	09BUN0	09H3P4	08C0A3	09Y3S1	09RNP7	07XIA0	082TR1	096BD8	0983M9	0964R1	09PR58	098PL8	07XLJ7	07XW87	08N608	07YI02	09Q8W3	09Q815	083655	08Y5G9	08TYM8	08YTB0	09FTE2	08LGM8	07XWU3	08YAI0	09XD61	07VU54	070241	08T262	08PL37	08P9D0	08UR52	08XAP8	08Z6R5	083ML7	099973	09X771	041461	083GPD	037435	081937	081058	09X303	09V164	004546	08R24	0882W4	088A12	08TR12	095W16	090220
--------------------	--------------------	-------------------	-------------------	--------------------	--------------------	--------	--------	--------	--------	--------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	-------	--------	--------	--------	--------	--------

90 42 38.2 492 2 Q8GN65  
 91 42 38.2 525 10 Q9AT04  
 92 42 38.2 588 10 Q7XM07  
 93 42 38.2 604 16 Q8FUM3  
 94 42 38.2 620 10 Q7XW15  
 95 42 38.2 637 5 Q96Q08  
 96 42 38.2 651 5 Q20710  
 97 42 38.2 806 10 Q7XS18  
 98 42 38.2 999 10 Q7XUY0  
 99 42 38.2 1078 10 Q8LIX9  
 100 42 38.2 1094 10 Q7XFP7

## ALIGNMENTS

## RESULT 1

ID Q59983 PRELIMINARY; PRT; 1590 AA.  
 AC Q59983;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5).  
 GN GTFI.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM2176;  
 RX MEDLINE=9416405; PubMed=8312602;  
 RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
 RT "MNA sequence of the glucosyltransferase gene of serotype d  
 Streptococcus sobrinus";  
 RL DNA Seq. 4:19-27(1993).  
 DR EMBL; D13858; BAA02976.1; -.  
 DR PIR; A39841; A39841.  
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; Glyco\_hydro.70.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF01473; CM binding.1; 13.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 KW Glycosyltransferase; Signal; Transferase.  
 FT SIGNAL 1  
 FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.  
 FT SIGNAL 38 POTENTIAL.  
 SQ SEQUENCE 1590 AA; 175955 MM; C3C83A57CF3C2B0E CRC64;  
 Query Match 100.0%; Score 110; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-08;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSSFPARHDSVQDIIRDTI 22  
 DB 548 VPSSFPARHDSVQDIIRDTI 569

## RESULT 2

ID Q55263 PRELIMINARY; PRT; 1590 AA.  
 AC Q55263;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE GTF-I  
 GN GLUCOSYLTRANSFERASE.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33478;  
 RA Sato S.;  
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase  
 produced from Streptococcus sobrinus ATCC 33478.";  
 RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
 DR EMBL; D63570; BAA09792.1; -.  
 DR PIR; A39841; A39841.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; Glyco\_hydro.70.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF01473; CM binding.1; 13.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 SQ SEQUENCE 1590 AA; 176058 MM; 9DF7A3F2C6E4FD43 CRC64;  
 Query Match 100.0%; Score 110; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-08;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSSFPARHDSVQDIIRDTI 22  
 DB 548 VPSSFPARHDSVQDIIRDTI 569

## RESULT 3

ID Q48756 PRELIMINARY; PRT; 1290 AA.  
 AC Q48756;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Dextranucrase.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B1299;  
 RX MEDLINE=97136686; PubMed=8982063;  
 RA Monchois V., Willemsot R.M., Renaud-Simeon M., Croux C., Monsan P.;  
 RT "Cloning and sequencing of a gene coding for a novel dextranucrase (1-  
 6) from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
 6) and alpha (1-3) linkages.";  
 RL Gene 182:23-32(1996).  
 DR EMBL; U38181; ABA0875.1; -.  
 DR PIR; JC3473; JC3473.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF01473; CM binding.1; 9.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 SQ SEQUENCE 1290 AA; 145590 MM; 3555C2E96B749FMA CRC64;  
 Query Match 77.3%; Score 85; DB 2; Length 1290;  
 Best Local Similarity 81.0%; Pred. No. 0.00018;  
 Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PYSFPARHDSVQDIIRDTI 22  
 DB 388 PYSFPARHDSVQDIIRDTI 408

## RESULT 4

ID Q84CM4 PRELIMINARY; PRT; 1330 AA.  
 AC Q84CM4;  
 DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Dextranucrase Derr (EC 2.4.1.5).  
 GN DSR.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.

```
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1501;
RA Kim C.H., Moon J.O., Jang E.K.;
RT "Gene encoding a dextranucrase (DsrR) in Leuconostoc mesenteroides
   NRRL B-1501."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR GO:0016757; P:glucan biosynthesis; IEA.
DR GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding_70.
DR InterPro: IPR00318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding_1; 13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1330 AA; 148663 MW; D945CBB36CF75797 CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1330;
Best Local Similarity 68.2%; Pred. No. 0.00055;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Cy 1 VPSSYPARADSEVQDIIRDI 22
Db 456 IPNYSFVRAHDSVQTVIAQII 477

RESULT 5
091466 PRELIMINARY; PRT; 1477 AA.
AC 091466;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN DSRC.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1355;
RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,
   Willemot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding altermannucrase, a sucrose
   RT glycosyltransferase from Leuconostoc mesenteroides NRRL B-1355."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ250172; CAB76565.1; -.
DR GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding_70.
DR InterPro: IPR00318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding_1; 13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1477 AA; 164887 MW; E6F5710DEDFC831 CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1477;
Best Local Similarity 68.2%; Pred. No. 0.00062;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Cy 1 VPSSYPARADSEVQDIIRDI 22
Db 603 IPNYSFVRAHDSVQTVIAQII 624

RESULT 6
09EZH5 PRELIMINARY; PRT; 1508 AA.
AC 09EZH5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE Dextranucrase DsrB742.
GN DSRB742.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-742CB;
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;
RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF294469; AAG38021.1; -.
DR GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding_70.
DR InterPro: IPR00318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding_1; 13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87ABA4F3A CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1508;
Best Local Similarity 68.2%; Pred. No. 0.00064;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Cy 1 VPSSYPARADSEVQDIIRDI 22
Db 634 IPNYSFVRAHDSVQTVIAQII 655

RESULT 7
052224 PRELIMINARY; PRT; 1508 AA.
AC 052224;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase (EC 2.4.1.5).
GN DSRB.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1299;
RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;
RT "Cloning and sequencing of a gene coding for an extracellular
   RT dextranucrase (DSRB) from Leuconostoc mesenteroides NRRL B-1299
   RT synthesizing only a a(1-6) glucan."
RL FEMS Microbiol. Lett. 0:0-0(1998).
DR EMBL: AF030129; AA895453.1; -.
DR PIR: T31098; T31098.
DR GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding_70.
DR InterPro: IPR00318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding_1; 13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1508 AA; 168511 MW; E70CEB57A70D1F0 CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1508;
Best Local Similarity 68.2%; Pred. No. 0.00064;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Cy 1 VPSSYPARADSEVQDIIRDI 22
Db 634 IPNYSFVRAHDSVQTVIAQII 655

RESULT 8
09LCJ7 PRELIMINARY; PRT; 1016 AA.
AC 09LCJ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DR Dextranucrase.  
 GN DSR.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=NRRL B-512F;  
 RX MEDLINE=20169623; PubMed=10705445;  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc mesenteroides NRRL B-512F."  
 RL Biosci. Biotechnol. Biochem. 64:29-38 (2000).  
 DR EMBL; AB020020; BAA90527.1; -.  
 DR HSP; P06278; LV05.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 SQ SEQUENCE 1016 AA; 110344 MW; 8896FDE13CCB47 CRC64;

Query Match 73.6%; Score 81; DB 2; Length 1016;  
 Best Local Similarity 71.4%; Pred. No. 0.00067;  
 Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPSYFARAHDSVQDIIRDI 22  
 Db 625 IPNYSFVARAHDSVQTVIAQIV 645

RESULT 9  
 Q9ZAR4 PRELIMINARY; PRT; 1527 AA.  
 AC Q9ZAR4;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DE Dextranucrase.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=NRRL B-512-F;  
 RX Bhatnagar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from Leuconostoc mesenteroides NRRL B-512F."  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U81374; AAD10952.1; -.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF01473; CM binding.1; 12.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 SQ SEQUENCE 1527 AA; 169708 MW; 1DFAFA237C743398 CRC64;

Query Match 73.6%; Score 81; DB 2; Length 1527;  
 Best Local Similarity 63.6%; Pred. No. 0.00093;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYFARAHDSVQDIIRDI 22  
 Db 652 IPNYSFVARAHDSVQTVIAQIV 673

RESULT 10  
 Q8KRE1 PRELIMINARY; PRT; 1527 AA.  
 AC Q8KRE1;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DR Dextranucrase Dard (EC 2.4.1.5).  
 GN DSRD.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Neubauer H., Bauche A., Mollet B.;  
 RT "Isolation and characterization of the dextranucrase Dard of Leuconostoc mesenteroides Lcc4."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AX013284; AAG61158.1; -.  
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . . IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF01473; CM binding.1; 12.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 DR Pfam; PF02324; Glycosyltransferase.  
 DR Transferrase; Glycosyltransferase.  
 SQ SEQUENCE 1527 AA; 169835 MW; P9D0DBE220BD89668 CRC64;

Query Match 73.6%; Score 81; DB 2; Length 1527;  
 Best Local Similarity 63.6%; Pred. No. 0.00093;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYFARAHDSVQDIIRDI 22  
 Db 652 IPNYSFVARAHDSVQTVIAQIV 673

RESULT 11  
 Q8KZL5 PRELIMINARY; PRT; 1554 AA.  
 AC Q8KZL5;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, last sequence update)  
 DT 01-JUN-2003 (Tremblrel. 24, last annotation update)  
 DE Glucosyltransferase.  
 GN GTFU.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21958684; PubMed=11960691;  
 RA Hanada N., Fukushima K., Nomura Y., Sempuku H., Hayakawa M.,  
 RA Mukasa H., Shiroza T., Abiko Y.;  
 RT "Cloning and nucleotide sequence analysis of the Streptococcus sobrinus gtfU gene that produces a highly branched water-soluble glucan."  
 RL Biochim. Biophys. Acta 1570:75-79 (2002).  
 DR EMBL; AB089438; BAC07265.1; -.  
 DR GO; GO:0016740; P:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro.70.  
 DR Pfam; PF01473; CM binding.1; 14.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 DR Pfam; PF02324; Glyco\_hydro.70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1554 AA; 171676 MW; 6981BC1DAE24A73 CRC64;

Query Match 72.7%; Score 80; DB 2; Length 1554;  
 Best Local Similarity 63.6%; Pred. No. 0.0014;  
 Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYFARAHDSVQDIIRDI 22  
 Db 557 IPTYSFVARAHDSVQTVIAQIV 578

```

RESULT 12
09WKJ5 PRELIMINARY; PRT: 1512 AA.
AC 09WKJ5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE GTP-S.
GN GTP.
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=HS-6;
RC Inoue M., Fukui K., Miyagi A.;
RT "S. cricetus glucosyltransferase (gifs and gftf) genes.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; C: binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; C: binding_1; 11.
DR Pfam; PF03224; Glyco_hydro_70; 1.
DR Plasmid.
KW SEQUENCE 1512 AA; 167145 MW; 4C03D9C8C601FC14 CRC64;
SQ

Query Match 69.1%; Score 76; DB 2; Length 1512;
Best Local Similarity 71.4%; Pred. No. 0.0057;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 PSYFARAHSEVQDIIRDI 22
Db 560 PSYFARAHSEVQTVIAQII 560

RESULT 13
08G9Q2 PRELIMINARY; PRT: 2835 AA.
AC 08G9Q2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DR Dextranucrase (RC 2.4.1.5) (Fragment).
GN DSR.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22231661; PubMed=12270834;
RA Mondan P., Willemot R.M.;
RA Bozonnet S., Dolé-Lafaygue M., Fabre E., Pizut S., Remaud-Simeon M.,
RT "Molecular characterisation of DSR-E, an alpha-1,2 linkage
RT synthesising dextranucrase with two catalytic domains.";
RL J. Bacteriol. 184:5753-5761(2002).
DR EMBL; AJ430204; CAD22883.1; -.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; C: binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; C: binding_1; 20.
DR Pfam; PF03224; Glyco_hydro_70; 2.
DR Transferase; Glycosyltransferase.
KW NON TER
FT SEQUENCE 2835 AA; 313264 MW; D03262CDD735399D CRC64;
SQ

Query Match 68.2%; Score 75; DB 2; Length 2835;
Best Local Similarity 66.7%; Pred. No. 0.016;
Matches 14; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 2 PSYFARAHSEVQDIIRDI 22
Db 629 PSYFARAHSEVQTVIAQII 649

RESULT 14
09LCH3 PRELIMINARY; PRT: 1575 AA.
AC 09LCH3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DR Glucosyltransferase.
GN GTPR.
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC10557;
RX MEDLINE=20231779; PubMed=10768934;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RT encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; C: binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; C: binding_1; 16.
DR Pfam; PF03224; Glyco_hydro_70; 1.
DR Transferase.
KW SEQUENCE 1575 AA; 176792 MW; 772A26B4D7C2B543 CRC64;
SQ

Query Match 65.5%; Score 72; DB 2; Length 1575;
Best Local Similarity 70.0%; Pred. No. 0.026;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 PSYFARAHSEVQDIIRDI 22
Db 619 PSYFARAHSEVQTVIAQII 638

RESULT 15
054178 PRELIMINARY; PRT: 1577 AA.
AC 054178;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DR Glucosyltransferase.
GN GTPG.
OS Streptococcus gordonii Challis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=29390;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
RT phase variants.";
RL Dev. Biol. Stand. 85:309-314(1995).
RN [2]
RP SEQUENCE OF 1-96 FROM N.A.
RA STRAIN=CHALLIS;
RX MEDLINE=92276337; PubMed=1534326;
RA Sulavik M.C., Tardif G., Clewell D.B.;
RT "Identification of a gene, rgg, which regulates expression of

```





```
RT coding for primer-independent glucosyltransferases."
RL Infect. Immun. 63:609-621(1995).
DR EMBL; J35495; AAC41412.1; -.
DR PIR; J30857; T30857.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Signal; Transferase.
KW SIGNAL
FT CHAIN 1 35 POTENTIAL.
FT SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;
SQ
Query Match 63.6%; Score 70; DB 2; Length 1449;
Best Local Similarity 70.0%; Pred. No. 0.049;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 3 SYSFARADSEVODIRDI 22
DB 609 NYAFVRAHDSVQSLIGQIT 628

RESULT 19
Q9WXC4 PRELIMINARY; PRT; 1338 AA.
AC Q9WXC4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTP-S.
GN GTPS.
OS Streptococcus criceti.
OC Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL "S.cricetus glucosyltransferase (gfs and gft) genes."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAAT7236.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Signal; Transferase.
KW PLASMID.
FT SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;
SQ
Query Match 62.7%; Score 69; DB 2; Length 1338;
Best Local Similarity 68.2%; Pred. No. 0.064;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 VPSYFARADSEVODIRDI 22
DB 509 VENVYFIRAHDSVQTRIKIT 530

RESULT 20
Q9RE05 PRELIMINARY; PRT; 2057 AA.
AC Q9RE05;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Alternansucrase (EC 2.4.1.140).
GN ASR.
OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1355;
RX MEDLINE=20080809; PubMed=10612736;
RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,
RA Willemot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355."
RL FEMS Microbiol. Lett. 182:81-85(2000).
DR EMBL; AJ250173; CAB55910.2; -.
DR GO; GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
FT SEQUENCE 2057 AA; 228967 MW; 62BCE9385D9A11BE CRC64;
SQ
Query Match 61.8%; Score 68; DB 2; Length 2057;
Best Local Similarity 63.2%; Pred. No. 0.15;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 VPSYFARADSEVODIR 19
DB 757 IPNYSFRAHDYDAQPIR 775

RESULT 21
Q00600 PRELIMINARY; PRT; 1518 AA.
AC Q00600;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Glucosyltransferase I (EC 2.4.1.5) (GTF) (Dextranucrase) (Sucrose 6-
DE glucosyltransferase).
GN GTFI.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least two
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975."
RL J. Gen. Microbiol. 137:2577-2593(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO
CC PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF
CC THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL) (N) = D-
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- DISEASE: DENTAL CARIES.
DR EMBL; Z11873; CAA79900.1; -.
DR EMBL; M64111; AAA26896.1; -.
DR PIR; A44811; A44811.
DR GO; GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferrase; Glycosyltransferase; Repeat; Dental caries.
FT DOMAIN 1307 1482 6 DIRECT REPEATS.
FT REPEAT 1307 1358 REPEAT 1.
FT REPEAT 1339 1352 REPEAT 2.
```

FT REPEAT 1372 1403 REPEAT 3.  
 FT REPEAT 1404 1417 REPEAT 4.  
 FT REPEAT 1437 1468 REPEAT 5.  
 FT REPEAT 1469 1482 REPEAT 6.  
 SQ SEQUENCE 1518 AA; 167730 MW; DAAAI717098B59A CRC64;

Query Match 60.9%; Score 67; DB 2; Length 1518;  
 Best Local Similarity 65.0%; Pred. No. 0.15;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDIIRDTI 22  
 ID 604 NYVFIKADHNNVDIIRDTI 623

RESULT 22  
 OS5265 PRELIMINARY; PRT; 1577 AA.

AC Q55265; PRELIMINARY; PRT; 1577 AA.  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, last annotation update)  
 DE Glucosyltransferase precursor.  
 GN GTFM.  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1304;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9512197; PubMed=7822030;  
 RA Simpson C.L., Giffard P.M., Jacques N.A.;  
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
 coding for primer-independent glucosyltransferases.";  
 RL Infect. Immun. 63:609-621(1995).  
 DR EMBL; L35928; AAC41413.1; -.  
 DR PIR; T30858; T30858.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR004829; Carbohydrate antigen.  
 DR InterPro; IPR002479; CM binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CM binding\_1; 10.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR ProDom; PD153432; Surface\_antigen; 1.  
 DR Signal; Transferase.

FT SIGNAL 1 POTENTIAL.  
 FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.  
 SQ SEQUENCE 1577 AA; 175290 MW; 3EPB898A7D3A7BF3 CRC64;

Query Match 58.2%; Score 64; DB 2; Length 1577;  
 Best Local Similarity 60.0%; Pred. No. 0.48;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDIIRDTI 22  
 ID 661 NYVFIKADHSEVQAVLANIT 680

RESULT 23  
 OS5265 PRELIMINARY; PRT; 108 AA.

AC P74028; PRELIMINARY; PRT; 108 AA.  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, last annotation update)  
 DE Hypothetical protein s111219.  
 GN SLI1219.  
 OS Synecchocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecchocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,  
 RA Miyajima N., Hirohata M., Sugitani M., Saito S., Kimura T.,  
 RA Hosouchi T., Matsuo A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL; D90911; BAA18101.1; -.  
 DR PIR; S75540; S75540.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 108 AA; 12030 MW; 5895770326CDBF CRC64;

Query Match 46.4%; Score 51; DB 16; Length 108;  
 Best Local Similarity 38.9%; Pred. No. 3;  
 Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 5 SEFARAHSEVQDIIRDTI 22  
 ID 34 NMAHARHSDLDVQDEIL 51

RESULT 24  
 OS21X9 PRELIMINARY; PRT; 93 AA.

AC Q921X9; PRELIMINARY; PRT; 93 AA.  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N40.  
 RA Feng S., Hodzic E., Barthold S.W.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF006036; AAD01254.1; -.  
 KM Hypothetical protein.  
 FT NON TER 1

SQ SEQUENCE 93 AA; 10129 MW; 58FDSDB8404FA6A CRC64;  
 Query Match 45.5%; Score 50; DB 2; Length 93;  
 Best Local Similarity 52.3%; Pred. No. 3.6;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 PSFSPRAHSSALIELI 18  
 ID 70 PSFSPRAHSSALIELI 86

RESULT 25  
 OS8026 PRELIMINARY; PRT; 1554 AA.

AC Q8026; PRELIMINARY; PRT; 1554 AA.  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)  
 DE MYO2.  
 OS Cryptococcus neoformans var. grubii.  
 CC Basidiomycota; Fungi; Basidiomycetes; Hymenomycetes; Heterobasidiomycetes;  
 CC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OX NCBI\_TaxID=178876;  
 RN [1]  
 RP SEQUENCE OF 940-1554 FROM N.A.

RC STRAIN=H99;  
 RC MEDLINE=20570501; PubMed=11121047;  
 RX Lengele K.B., Wang P., Cox G.M., Perfect J.R., Heitman J.;  
 RA "Identification of the MATa mating-type locus of Cryptococcus  
 RT neoformans reveals a serotype A MATa strain thought to have been

```

RT extinct."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14455-14460(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H99;
RX MEDLINE=2243086; PubMed=12455690;
RA Lengeler K.B., Fox D.S., Fraser J.A., Allen A., Forrester K.,
RA Dieterich P.S., Heitman J.,
RT "Mating-type locus of Cryptococcus neoformans: a step in the evolution
RT of sex chromosomes."
RL Eukaryot. Cell 1:704-718(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H99;
RA Wang P., Heitman J.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=H99;
RA Lengeler K.B., Fox D.S., Fraser J.A., Allen A., Forrester K.,
RA Dieterich P.S., Heitman J.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF542529; AAN75169.1; -.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 1554 AA; 174915 MW; ED0C4545648835 CRC64;

Query Match 45.5%; Score 50; DB 3; Length 1554;
Best Local Similarity 47.4%; Pred. No. 78;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSHVQDIIR 19
DB 1253 VPGYDFSGHSDSDWGRGIR 1271

RESULT 26
Q8J0W2 PRELIMINARY; PRT; 1567 AA.
AC Q8J0W2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MYO2.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellales; Tremellaceae; Tremellales; Filobasidiella.
OC NCBI_Taxid=40410;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=JEC21;
RC MEDLINE=22343086; PubMed=12455690;
RA Lengeler K.B., Fox D.S., Fraser J.A., Allen A., Forrester K.,
RA Dieterich P.S., Heitman J.;
RT "Mating-type locus of Cryptococcus neoformans: a step in the evolution
RT of sex chromosomes."
RL Eukaryot. Cell 1:704-718(2002).
DR EMBL; AF542531; AAN75723.1; -.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR002710; DIL.

```

```

DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 1567 AA; 176363 MW; D875A36BB8BEBAF2 CRC64;

Query Match 45.5%; Score 50; DB 3; Length 1567;
Best Local Similarity 47.4%; Pred. No. 79;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPSYSPARAHDSHVQDIIR 19
DB 1266 VPGYDFSGHSDSDWGRGIR 1284

RESULT 27
Q8A338 PRELIMINARY; PRT; 303 AA.
AC Q8A338;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative oxidoreductase.
GN BT3117.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_Taxid=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076(2003).
DR EMBL; AB016939; AAO78223.1; -.
DR GO; GO:0008677; F:2-dehydropanoate 2-reductase activity; IEA.
DR GO; GO:0000036; F:acyl carrier activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0006206; P:pyrimidine base metabolism; IEA.
DR InterPro; IPR008927; 6DGDH_C-like.
DR InterPro; IPR003231; Acyl_carrier.
DR InterPro; IPR003710; AbpA.
DR Pfam; PF02558; AbpA; 1.
DR ProDom; PD000887; Acyl_carrier; 1.
DR TIGRfams; TIGR00745; abpA_pamB; 1.
KW Complete proteome.
SQ SEQUENCE 303 AA; 32667 MW; 72D5D7250673271E CRC64;

Query Match 42.7%; Score 47; DB 16; Length 303;
Best Local Similarity 50.0%; Pred. No. 39;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 7 ARAHDSHVQDIIRDI 22
DB 265 ARGHSEIRIGLLFDWI 280

RESULT 28
Q8R6U3 PRELIMINARY; PRT; 449 AA.
AC Q8R6U3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Membrane proteins related to metalloendopeptidases.
GN NLPD6 OR TTE2691.

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:07:31 ; Search time 14 Seconds  
(without alignments)  
81.127 Million cell updates/sec

Title: US-09-290-049A-19  
Perfect score: 110  
Sequence: 1 VPSYFARAHDSVQDIIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCUTS COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	98	89.1	1375 3 US-09-210-361-4	Sequence 4, Appli
2	98	89.1	1375 4 US-09-740-274-4	Sequence 4, Appli
3	97	88.2	1475 3 US-09-007-999-2	Sequence 2, Appli
4	97	88.2	1475 3 US-09-210-361-2	Sequence 2, Appli
5	97	88.2	1475 4 US-09-740-274-2	Sequence 2, Appli
6	97	88.2	1475 4 US-09-604-957-5	Sequence 5, Appli
7	68	61.8	535 4 US-09-604-957-7	Sequence 7, Appli
8	68	61.8	1278 4 US-09-604-957-3	Sequence 3, Appli
9	68	61.8	2057 4 US-09-499-203-2	Sequence 2, Appli
10	64	58.2	545 4 US-09-604-957-4	Sequence 4, Appli
11	64	58.2	1430 3 US-09-008-172-2	Sequence 2, Appli
12	64	58.2	1430 3 US-09-210-361-6	Sequence 6, Appli
13	64	58.2	1430 3 US-09-740-274-6	Sequence 6, Appli
14	64	58.2	1577 2 US-08-793-824-2	Sequence 2, Appli
15	63	57.3	584 4 US-09-604-957-6	Sequence 6, Appli
16	46.5	42.3	501 4 US-09-134-001C-4115	Sequence 4115, Ap
17	46	41.8	309 4 US-09-345-473E-37	Sequence 37, Appli
18	45.5	41.4	484 3 US-08-913-578-2	Sequence 2, Appli
19	45.5	41.4	484 3 US-08-785-427-2	Sequence 2, Appli
20	44	40.0	565 4 US-09-107-532A-4217	Sequence 4217, Ap
21	43	39.1	2627 2 US-08-751-189-3	Sequence 3, Appli
22	43	39.1	2627 2 US-09-060-836-3	Sequence 3, Appli
23	43	39.1	2627 3 US-09-184-445-3	Sequence 3, Appli
24	42	38.2	196 2 US-08-684-024-9	Sequence 2, Appli
25	42	38.2	196 2 US-08-684-024-9	Sequence 9, Appli
26	42	38.2	196 3 US-09-145-868-2	Sequence 2, Appli
27	42	38.2	196 3 US-09-145-868-9	Sequence 9, Appli

28	42	38.2	392 4 US-09-424-978B-39	Sequence 39, Appli
29	41	37.3	639 2 US-08-557-309B-37	Sequence 37, Appli
30	41	37.3	639 3 US-08-834-306-37	Sequence 37, Appli
31	41	37.3	639 3 US-08-993-674A-37	Sequence 37, Appli
32	41	37.3	639 3 US-09-256-976-37	Sequence 2, Appli
33	40.5	36.8	480 2 US-08-844-153-2	Sequence 2, Appli
34	40.5	36.8	480 2 US-08-962-203-2	Sequence 2, Appli
35	40.5	36.8	480 3 US-09-282-125A-2	Sequence 2, Appli
36	40.5	36.8	480 4 US-09-273-142-2	Sequence 2, Appli
37	40	36.4	75 4 US-09-107-532A-7020	Sequence 7020, Ap
38	40	36.4	252 4 US-08-858-207A-325	Sequence 325, App
39	40	36.4	265 4 US-09-540-236-2661	Sequence 2661, Ap
40	40	36.4	340 4 US-09-328-352-7834	Sequence 7834, Ap
41	40	36.4	439 4 US-09-489-039A-8498	Sequence 8498, Ap
42	40	36.4	639 4 US-09-268-347-38	Sequence 38, Appli
43	40	36.4	639 4 US-09-509-814A-4	Sequence 4, Appli
44	40	36.4	659 4 US-09-268-347-46	Sequence 46, Appli
45	40	36.4	660 4 US-09-268-347-45	Sequence 45, Appli
46	40	36.4	1068 3 US-08-390-874C-11	Sequence 11, Appli
47	40	36.4	1068 4 US-09-265-772-11	Sequence 11, Appli
48	40	36.4	1069 2 US-08-162-081B-37	Sequence 37, Appli
49	40	36.4	1069 2 US-08-780-872-37	Sequence 37, Appli
50	40	36.4	1069 3 US-09-085-957-37	Sequence 36, Appli
51	40	36.4	1080 2 US-08-162-081B-36	Sequence 36, Appli
52	40	36.4	1080 2 US-08-780-872-36	Sequence 36, Appli
53	40	36.4	1080 2 US-09-085-957-36	Sequence 36, Appli
54	39	35.5	184 1 US-08-353-550-2	Sequence 2, Appli
55	39	35.5	184 2 US-08-551-687-2	Sequence 2, Appli
56	39	35.5	236 3 US-08-935-263-2	Sequence 2, Appli
57	39	35.5	236 4 US-09-594-185-2	Sequence 2, Appli
58	39	35.5	333 4 US-09-252-991A-33113	Sequence 33113, A
59	39	35.5	334 4 US-09-489-039A-8909	Sequence 8909, Ap
60	39	35.5	409 4 US-09-134-000C-6168	Sequence 6168, Ap
61	39	35.5	409 4 US-08-353-550-1	Sequence 1, Appli
62	39	35.5	409 4 US-08-551-687-1	Sequence 1, Appli
63	39	35.5	431 3 US-08-807-342D-5	Sequence 5, Appli
64	39	35.5	469 1 US-08-353-550-6	Sequence 6, Appli
65	39	35.5	469 2 US-08-551-687-6	Sequence 6, Appli
66	39	35.5	551 4 US-09-252-991A-25323	Sequence 25323, A
67	39	35.5	933 1 US-08-370-193A-8	Sequence 8, Appli
68	39	35.5	933 4 US-09-271-438A-9	Sequence 9, Appli
69	39	35.5	933 4 US-10-078-107-5	Sequence 5, Appli
70	39	35.5	933 4 US-10-077-751-5	Sequence 5, Appli
71	39	35.5	935 4 US-09-271-438A-3	Sequence 3, Appli
72	39	35.5	935 4 US-09-271-438A-8	Sequence 8, Appli
73	39	35.5	935 4 US-10-078-107-1	Sequence 1, Appli
74	39	35.5	935 4 US-10-077-751-1	Sequence 1, Appli
75	39	35.5	938 4 US-09-489-039A-13504	Sequence 13504, A
76	39	35.5	1365 4 US-09-376-330-18	Sequence 18, Appli
77	39	35.5	1365 6 5194600-4	Sequence 21269, A
78	39	35.5	143 3 US-08-679-006-2	Sequence 2, Appli
79	38	34.5	90 4 US-09-388-395A-35	Sequence 35, Appli
80	38	34.5	90 4 US-09-887-586A-35	Sequence 35, Appli
81	38	34.5	90 4 US-09-895-752-35	Sequence 35, Appli
82	38	34.5	90 4 US-09-903-012B-35	Sequence 35, Appli
83	38	34.5	90 4 US-09-900-797-35	Sequence 35, Appli
84	38	34.5	128 4 US-09-134-001C-5564	Sequence 9, Appli
85	38	34.5	128 4 US-09-511-024A-9	Sequence 7052, Ap
86	38	34.5	191 4 US-09-328-352-7052	Sequence 9032, Ap
87	38	34.5	235 4 US-09-489-039A-9032	Sequence 18759, A
88	38	34.5	321 4 US-09-252-991A-18759	Sequence 8, Appli
89	38	34.5	348 3 US-09-188-955A-8	Sequence 8, Appli
90	38	34.5	348 4 US-09-654-531-8	Sequence 8, Appli
91	38	34.5	348 4 US-10-072-152-8	Sequence 8, Appli
92	38	34.5	372 4 US-09-489-039A-12346	Sequence 12346, A
93	38	34.5	402 4 US-09-029-755C-5	Sequence 7, Appli
94	38	34.5	402 5 PCT-US94-01321-72	Sequence 21718, A
95	38	34.5	440 4 US-09-252-991A-25869	Sequence 25869, A
96	38	34.5	450 4 US-09-252-991A-4693	Sequence 4693, A
97	38	34.5	501 4 US-09-134-000C-13795	Sequence 13795, A
98	38	34.5	557 4 US-09-489-039A-13795	Sequence 28, Appli
99	38	34.5	578 4 US-09-857-556A-28	
100	38	34.5		

## ALIGNMENTS

```
RESULT 1
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-4

Query Match      89.1%; Score 98; DB 3; Length 1375;
Best Local Similarity 86.4%; Pred. No. 1e-07;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 VPSYSPRAHDSVQDIIRDI 22
      ||||| ||||| ||||| |||||
DB      578 VPSYSPRAHDSVQDIIRNI 599

RESULT 2
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/740,274
; EARLIER FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
```

```
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-4

Query Match      89.1%; Score 98; DB 4; Length 1375;
Best Local Similarity 86.4%; Pred. No. 1e-07;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 VPSYSPRAHDSVQDIIRDI 22
      ||||| ||||| ||||| |||||
DB      578 VPSYSPRAHDSVQDIIRNI 599

RESULT 3
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match      88.2%; Score 97; DB 3; Length 1475;
Best Local Similarity 86.4%; Pred. No. 1.6e-07;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 VPSYSPRAHDSVQDIIRDI 22
      ||||| ||||| ||||| |||||
DB      552 VPSYSPRAHDSVQDIIRNI 573

RESULT 4
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
```

Query Match 88.2%; Score 97; DB 3; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 1.6e-07;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYSPARADSEVQDIIRDI 22  
Db 552 VPSYSPARADSEVQDIIRDI 573

## RESULT 5

US-09-740-274-2  
; Sequence 2, Application US/09740274  
; Patent No. 6465203  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLS, SCOTT E.  
; TITLE OF INVENTION: GLUCAN-CONTAINING COMPOSITIONS AND PAPER  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 88.2%; Score 97; DB 4; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 1.6e-07;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYSPARADSEVQDIIRDI 22  
Db 552 VPSYSPARADSEVQDIIRDI 573

## RESULT 6

US-09-604-957-5  
; Sequence 5, Application US/09604957  
; Patent No. 6486314  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Qy 1 VPSYSPARADSEVQDIIRDI 21  
Db 620 IPNYSFVARHNNNSQDIQNAIRDV 644

Query Match 73.6%; Score 81; DB 4; Length 523;  
Best Local Similarity 63.6%; Pred. No. 2.2e-05;  
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPSYSPARADSEVQDIIRDI 22  
Db 146 IPNYSFVARHNNNSQDIQNAIRDV 167

## RESULT 7

US-09-604-957-7  
; Sequence 7, Application US/09604957  
; Patent No. 6486314  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-09-604-957-7

Query Match 61.8%; Score 68; DB 4; Length 535;  
Best Local Similarity 48.0%; Pred. No. 0.0033;  
Matches 12; Conservative 7; Mismatches 2; Indels 4; Gaps 1;

Qy 1 VPSYSPARADSEVQDIIRDI 21  
Db 144 IPNYSFVARHNNNSQDIQNAIRDV 168

## RESULT 8

US-09-604-957-3  
; Sequence 3, Application US/09604957  
; Patent No. 6486314  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1278  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-09-604-957-3

Query Match 61.8%; Score 68; DB 4; Length 1278;  
Best Local Similarity 48.0%; Pred. No. 0.0033;  
Matches 12; Conservative 7; Mismatches 2; Indels 4; Gaps 1;

Qy 1 VPSYSPARADSEVQDIIRDI 21  
Db 620 IPNYSFVARHNNNSQDIQNAIRDV 644



RESULT 9  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; Patent No. 6570065  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-1969  
; CURRENT APPLICATION NUMBER: US/09/499,203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2057  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 61.8%; Score 68; DB 4; Length 2057;  
Best Local Similarity 63.2%; Pred. No. 0.016;  
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 VPSYFRAHDSVQDIIR 19  
DB 757 NYIFRAHDSVQDIIR 775

RESULT 10  
US-09-604-957-4  
; Sequence 4, Application US/09604957  
; Patent No. 6486314  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUTI, HAKIM  
; APPLICANT: LIEB, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 58.2%; Score 64; DB 4; Length 545;  
Best Local Similarity 65.0%; Pred. No. 0.016;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 SYSFARHDSVQDIIR 22  
DB 156 NYIFRAHDSVQDIIR 175

RESULT 11  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172

; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 58.2%; Score 64; DB 3; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.049;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 SYSFARHDSVQDIIR 22  
DB 576 NYIFRAHDSVQDIIR 595

RESULT 12  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 09/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 58.2%; Score 64; DB 3; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.049;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 SYSFARHDSVQDIIR 22  
DB 576 NYIFRAHDSVQDIIR 595

RESULT 13  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; Patent No. 6465203  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999

PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 58.2%; Score 64; DB 4; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.049;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARADSEVODIIRII 22  
Db 576 NYIFRADSEVQTVIAKII 595

RESULT 14  
US-08-793-824-2  
Sequence 2, Application US/08793824  
Patent No. 5981838  
GENERAL INFORMATION:  
APPLICANT: Simpson, Christine Lynn  
APPLICANT: Giffard, Philip Morrison  
APPLICANT: Jacques, Nicholas Anthony  
TITLE OF INVENTION: Genetic Manipulation of Plants to  
TITLE OF INVENTION: Increase Stored Carbohydrates  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Griffith Hack & Co  
STREET: Level 8, 168 Walker Street  
CITY: No. 5981838th Sydney  
STATE: New South Wales  
COUNTRY: Australia  
ZIP: 2060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,824  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PM7643  
FILING DATE: 24-AUG-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 61 2 9957 5944  
TELEFAX: 61 2 957 6288  
TELEX: 26547  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1577 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 58.2%; Score 64; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.056;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARADSEVODIIRII 22  
Db 661 NYIFRADSEVQAVLANII 680

RESULT 15  
US-09-604-957-6  
Sequence 6, Application US/09604957  
Patent No. 6486314  
GENERAL INFORMATION:  
APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOUBERT  
APPLICANT: RAHADI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 4338  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 57.3%; Score 63; DB 4; Length 584;  
Best Local Similarity 61.1%; Pred. No. 0.025;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYFARADSEVODII 18  
Db 167 IPNYSFARADYDADEI 184

RESULT 16  
US-09-134-001C-4115  
Sequence 4115, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: CTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4115  
LENGTH: 501  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4115

Query Match 42.3%; Score 46.5; DB 4; Length 501;  
Best Local Similarity 45.5%; Pred. No. 12;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYFARA---HDESVODIIR 19  
Db 210 VPTNFAVADHYMDVIR 231

RESULT 17



CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4217:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 565 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...565  
SEQUENCE DESCRIPTION: SEQ ID NO: 4217:  
US-09-107-532A-4217

Query Match 40.0%; Score 44; DB 4; Length 565;  
Best Local Similarity 36.8%; Pred. No. 35;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 PSYSFARADSEVODIRD 20  
DB 480 PSIAFVAITHIELEIKN 498

RESULT 21  
US-08-751-189-3  
Sequence 3, Application US/08751189  
Patent No. 5919656  
GENERAL INFORMATION:  
APPLICANT: Harrington, Lea A.  
APPLICANT: Robinson, Murray O.  
TITLE OF INVENTION: No. 5919656e1 Genes Encoding Telomerase Protein  
TITLE OF INVENTION: 1  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen, Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/751,189  
FILING DATE: 15-NOV-1996

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oleksi, Nancy A.  
REGISTRATION NUMBER: 34,688  
REFERENCE/DOCKET NUMBER: A-433  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2627 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-751-189-3

Query Match 39.1%; Score 43; DB 2; Length 2627;  
Best Local Similarity 41.2%; Pred. No. 3.2e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYFARADSEVODI 17  
DB 205 MPYSLSIGREBERVDL 221

RESULT 22  
US-09-060-836-3  
Sequence 3, Application US/09060836  
Patent No. 5981707  
GENERAL INFORMATION:  
APPLICANT: Harrington, Lea A.  
APPLICANT: Robinson, Murray O.  
TITLE OF INVENTION: No. 5981707e1 Genes Encoding Telomerase Protein  
TITLE OF INVENTION: 1  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen, Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/060,836  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/751,189  
FILING DATE: 15-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Oleksi, Nancy A.  
REGISTRATION NUMBER: 34,688  
REFERENCE/DOCKET NUMBER: A-433  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2627 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-060-836-3

Query Match 39.1%; Score 43; DB 2; Length 2627;  
Best Local Similarity 41.2%; Pred. No. 3.2e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYFARADSEVODI 17  
DB 205 MPYSLSIGREBERVDL 221

RESULT 23  
US-09-184-445-3  
Sequence 3, Application US/09184445  
Patent No. 6174703  
GENERAL INFORMATION:  
APPLICANT: Harrington, Lea A.  
ATTORNEY/AGENT INFORMATION:  
APPLICANT: Robinson, Murray O.  
TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
TITLE OF INVENTION: 1  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Amgen, Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/184,445  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/751,189  
FILING DATE: 15-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Oleski, Nancy A.  
REGISTRATION NUMBER: 34,688  
REFERENCE/DOCKET NUMBER: A-433  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2627 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-184-445-3  
Query Match  
Best Local Similarity 39.1%; Score 43; DB 3; Length 2627;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 1 VPSYSPARAHSEVQDI 17  
Db 205 MPYSISLGEHEVEDL 221  
RESULT 24  
US-08-684-024-2  
Sequence 2, Application US/08684024  
Patent No. 5834298  
GENERAL INFORMATION:  
APPLICANT: Beneza, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/684,024  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-684-024-2  
Query Match  
Best Local Similarity 38.2%; Score 42; DB 2; Length 196;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 8 RAHDSVQDIIRDI 22  
Db 49 KTHDELDYIRKIL 63  
RESULT 25  
US-08-684-024-9  
Sequence 9, Application US/08684024  
Patent No. 5834298  
GENERAL INFORMATION:  
APPLICANT: Beneza, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684,024  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-684-024-9  
Query Match  
Best Local Similarity 38.2%; Score 42; DB 2; Length 196;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 8 RAHDSVQDIIRDI 22

Db : ||| : ||| :  
49 KTHDELKDYIRKIL 63

## RESULT 26

US-09-145-868-2  
Sequence 2, Application US/09145868  
Patent No. 6096522  
GENERAL INFORMATION:  
APPLICANT: Beneza, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-2

Query Match 38.2% Score 42; DB 3; Length 196;  
Best Local Similarity 46.7% Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
Cy 8 RAHSEVQDIIRDI 22  
Db 49 KTHDELKDYIRKIL 63

## RESULT 27

US-09-145-868-9  
Sequence 9, Application US/09145868  
Patent No. 6096522  
GENERAL INFORMATION:  
APPLICANT: Beneza, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-9

Query Match 38.2% Score 42; DB 3; Length 196;  
Best Local Similarity 46.7% Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
Cy 8 RAHSEVQDIIRDI 22  
Db 49 KTHDELKDYIRKIL 63

## RESULT 28

US-09-424-978B-39  
Sequence 39, Application US/09424978B  
Patent No. 666445  
GENERAL INFORMATION:  
APPLICANT: Falco, Saverio Carl  
APPLICANT: Allen, Stephen M.  
APPLICANT: Rafalski, J. Antoni  
APPLICANT: Hiltz, William D.  
APPLICANT: Kinney, Anthony J.  
APPLICANT: Abell, Lynne N.  
APPLICANT: Thorpe, Catherine J.  
TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes  
FILE REFERENCE: BB-1087  
CURRENT APPLICATION NUMBER: US/09/424,978B  
CURRENT FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: US 60/048,771  
PRIOR FILING DATE: 1997-06-06  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 39  
LENGTH: 392  
TYPE: PRT  
ORGANISM: Glycine max  
US-09-424-978B-39

Query Match 38.2% Score 42; DB 4; Length 392;  
Best Local Similarity 30.0% Pred. No. 49;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
Cy 1 VPSYFARHSEVQDIIRD 20  
Db 321 VDTYGTGKHKDKELIINIVKE 340

## RESULT 29

US-08-557-309B-37  
Sequence 37, Application US/08557309B  
Patent No. 5916572  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond L.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,309B  
FILING DATE: 14-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.422  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 682-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 639 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-557-309B-37

Query Match 37.3%; Score 41; DB 2; Length 639;  
Best Local Similarity 47.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDI 17  
DB 479 VPGWSEALHDAEFQOL 495

RESULT 30  
US-08-834-306-37  
Sequence 37, Application US/08834306  
Patent No. 6054135  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yahir A.W.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond L.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/834,306  
FILING DATE: 15-APR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.422C1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 682-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 639 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-834-306-37

Query Match 37.3%; Score 41; DB 3; Length 639;  
Best Local Similarity 47.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDI 17  
DB 479 VPGWSEALHDAEFQOL 495

Search completed: May 4, 2004, 09:14:13  
Job time : 15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2004, 09:12:41 ; Search time 35.333 Seconds  
(without alignments)  
172.590 Million cell updates/sec

Title: US-09-290-049A-19

Perfect score: 110

Sequence: 1 VPSSYSPARADSEVQDIIRDI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	89.1	1375	9	US-09-740-274-4
2	97	88.2	1475	9	US-09-740-274-2
3	81	73.6	522	9	US-09-995-749A-11
4	68	61.8	535	9	US-09-995-749A-13
5	68	61.8	584	9	US-09-995-749A-12
6	68	61.8	1781	9	US-09-995-749A-2
7	64	58.2	2057	15	US-10-417-280A-2
8	64	58.2	545	9	US-09-995-749A-10
9	64	58.2	1430	9	US-09-740-274-6
10	49	44.5	117	12	US-10-424-599-168435
11	48.5	44.1	465	12	US-10-282-122A-71700
12	47.5	43.2	481	9	US-09-815-242-5584
13	47.5	43.2	487	9	US-09-815-242-12456
14	47	42.7	166	12	US-10-424-599-264904
15	47	42.7	963	12	US-10-424-599-194106

16	46.5	42.3	484	12	US-10-282-122A-71074	Sequence 71074, A
17	46	41.8	115	12	US-10-424-599-267457	Sequence 267457, A
18	46	41.8	309	9	US-09-862-027-37	Sequence 37, Appl
19	46	41.8	490	12	US-10-276-774-2103	Sequence 2103, Ap
20	46	41.8	521	15	US-10-131-410-117	Sequence 117, App
21	46	41.8	564	15	US-10-131-410-180	Sequence 180, App
22	46	41.8	779	14	US-10-353-929-49	Sequence 49, Appl
23	46	41.8	1345	16	US-10-433-794-17	Sequence 17, Appl
24	45.5	41.4	484	12	US-10-282-122A-43803	Sequence 43803, A
25	45	40.9	54	12	US-10-424-599-232822	Sequence 232822, A
26	45	40.9	111	12	US-10-424-599-229462	Sequence 229462, A
27	45	40.9	171	12	US-10-424-599-184218	Sequence 184218, A
28	45	40.9	174	12	US-10-424-599-199858	Sequence 199858, A
29	45	40.9	255	9	US-09-895-828-454	Sequence 454, App
30	45	40.9	255	14	US-10-114-666-454	Sequence 454, App
31	45	40.9	262	9	US-09-895-828-457	Sequence 457, App
32	45	40.9	262	14	US-10-114-666-457	Sequence 457, App
33	45	40.9	462	15	US-10-369-493-12332	Sequence 12332, A
34	45	40.9	597	12	US-10-282-122A-76707	Sequence 76707, A
35	45	40.9	704	14	US-10-213-990-21	Sequence 21, Appl
36	44.5	40.5	386	15	US-10-012-697-1539	Sequence 1539, Ap
37	44.5	40.5	486	9	US-09-815-242-13455	Sequence 13455, A
38	44.5	40.5	486	12	US-10-282-122A-74199	Sequence 74199, A
39	44	40.0	170	12	US-10-424-599-159618	Sequence 159618, A
40	44	40.0	377	12	US-10-424-599-194041	Sequence 194041, A
41	44	40.0	423	12	US-10-425-114-65616	Sequence 65616, A
42	44	40.0	547	12	US-10-425-114-65616	Sequence 65616, A
43	43	39.1	79	12	US-10-424-599-284550	Sequence 284550, A
44	43	39.1	119	12	US-10-424-599-172541	Sequence 172541, A
45	43	39.1	211	12	US-10-425-114-50726	Sequence 50726, A
46	43	39.1	279	12	US-10-424-599-211056	Sequence 211056, A
47	43	39.1	362	12	US-10-424-599-158664	Sequence 158664, A
48	43	39.1	365	12	US-10-424-599-250160	Sequence 250160, A
49	43	39.1	377	12	US-10-425-114-53178	Sequence 53178, A
50	43	39.1	414	12	US-10-425-114-50726	Sequence 50726, A
51	43	39.1	479	15	US-10-369-493-20600	Sequence 20600, A
52	43	39.1	557	12	US-09-934-455-326	Sequence 326, App
53	43	39.1	557	12	US-10-412-699B-870	Sequence 870, App
54	43	39.1	557	12	US-10-302-267-198	Sequence 198, App
55	43	39.1	557	12	US-10-302-267-198	Sequence 198, App
56	43	39.1	557	15	US-10-374-780A-2598	Sequence 2598, App
57	43	39.1	2630	15	US-10-374-780A-2598	Sequence 2598, App
58	42.5	38.6	484	12	US-10-282-122A-64122	Sequence 64122, A
59	42.5	38.6	540	15	US-10-369-493-1936	Sequence 1936, Ap
60	42	38.2	66	12	US-10-424-599-252560	Sequence 252560, A
61	42	38.2	141	12	US-10-424-599-212103	Sequence 212103, A
62	42	38.2	145	12	US-10-424-599-151538	Sequence 151538, A
63	42	38.2	146	12	US-10-424-599-151538	Sequence 151538, A
64	42	38.2	150	12	US-10-424-599-223495	Sequence 223495, A
65	42	38.2	170	12	US-10-424-599-223495	Sequence 223495, A
66	42	38.2	185	12	US-10-425-114-61839	Sequence 61839, A
67	42	38.2	196	15	US-10-425-114-43588	Sequence 43588, A
68	42	38.2	305	12	US-10-369-493-22169	Sequence 22169, A
69	42	38.2	324	12	US-10-425-114-43785	Sequence 43785, A
70	42	38.2	333	12	US-10-425-114-44405	Sequence 44405, A
71	42	38.2	392	12	US-10-424-599-220046	Sequence 220046, A
72	42	38.2	393	12	US-10-424-599-220047	Sequence 220047, A
73	42	38.2	393	12	US-10-424-599-258837	Sequence 258837, A
74	42	38.2	395	12	US-10-425-114-44212	Sequence 44212, A
75	42	38.2	395	12	US-10-425-114-44212	Sequence 44212, A
76	42	38.2	413	12	US-10-425-114-44833	Sequence 44833, A
77	42	38.2	413	12	US-10-425-114-44833	Sequence 44833, A
78	42	38.2	416	12	US-10-425-114-43817	Sequence 43817, A
79	42	38.2	416	12	US-10-425-114-43817	Sequence 43817, A
80	42	38.2	420	12	US-10-425-114-45878	Sequence 45878, A
81	42	38.2	420	12	US-10-425-114-45878	Sequence 45878, A
82	42	38.2	421	12	US-10-425-114-55724	Sequence 55724, A
83	42	38.2	421	12	US-10-425-114-55724	Sequence 55724, A
84	42	38.2	421	12	US-10-425-114-55444	Sequence 55444, A
85	42	38.2	421	12	US-10-425-114-51917	Sequence 51917, A
86	42	38.2	422	12	US-10-425-114-51415	Sequence 51415, A
87	42	38.2	423	12	US-10-425-114-43754	Sequence 43754, A
88	42	38.2	423	12	US-10-425-114-44216	Sequence 44216, A



```
89 42 38.2 423 12 US-10-425-114-45712 Sequence 45712, A
90 42 38.2 423 12 US-10-425-114-45723 Sequence 45723, A
91 42 38.2 423 12 US-10-425-114-51430 Sequence 51430, A
92 42 38.2 423 12 US-10-425-114-53367 Sequence 53367, A
93 42 38.2 423 12 US-10-425-114-68219 Sequence 68219, A
94 42 38.2 423 12 US-10-425-114-71903 Sequence 71903, A
95 42 38.2 423 12 US-10-425-114-71916 Sequence 71916, A
96 42 38.2 423 12 US-10-425-114-71922 Sequence 71922, A
97 42 38.2 423 12 US-10-425-114-71923 Sequence 71923, A
98 42 38.2 423 12 US-10-425-114-71929 Sequence 71929, A
99 42 38.2 423 12 US-10-425-114-71954 Sequence 71954, A
100 42 38.2 423 12 US-10-425-114-71966 Sequence 71966, A
```

## ALIGNMENTS

```
RESULT 1
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRF
; ORGANISM: Streptococcus mutans
US-09-740-274-4

Query Match          89.1%; Score 98; DB 9; Length 1375;
Best Local Similarity 86.4%; Pred. No. 4,4e-07;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDIIRDI 22
Db 578 VPSYSPARAHSEVQDIIRDI 599

RESULT 2
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
```

```
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRF
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match          88.2%; Score 97; DB 9; Length 1475;
Best Local Similarity 86.4%; Pred. No. 7e-07;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDIIRDI 22
Db 552 VPSYSPARAHSEVQDIIRDI 573
```

```
RESULT 3
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. US2002015568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJCHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRF
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

Query Match          73.6%; Score 81; DB 9; Length 522;
Best Local Similarity 63.6%; Pred. No. 9.2e-05;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVQDIIRDI 22
Db 146 IPNYSFVARAHSEVQTVIAQIV 167

RESULT 4
US-09-995-749A-13
; Sequence 13, Application US/09995749A
; Patent No. US2002015568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJCHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
```

? SEQ ID NO 13  
 ? LENGTH: 535  
 ? TYPE: PRT  
 ? ORGANISM: *Lactobacillus reuteri*  
 US-09-995-749A-13

Query Match	61.8%	Score 68	DB 9	Length 535
Best Local Similarity	48.0%	Pred. No. 0.013		
Matches	12	Conservative	7	Mismatches 2
				Indels 4
				Gaps 1
QY	1	VPSSPARAHDS-----EYODIRIRDI	21	
		:     :     :     :		
Db	144	IPNTSFVRAHNDNSQDQIQNNAIRDV	168	

```

RESULT 5
US-09-995-749A-12
; Sequence 12, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL, SCHOTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHOUI, HAKIM
; APPLICANT: LEEB, ROBERT-JUN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 584
; TYPE: PRN
; ORGANISM: Leuconostoc mesenteroides
; US-09-995-749A-12

```

Query Match	61.8%	Score 68	DB 9	Length 584
Best Local Similarity	63.2%	Pred. NO. 0.015		
Matches 12	Conservative 3	Mismatches 4	Indels 0	Gaps 0

```
QY      1 VPSYSPARAHDSVEVDIIR 19
          :|:|:|:|:|:|:|:|:|
Db      167 IPNYSFVRADHYDAQDPIR 185
```

RESULT 6  
 US-09-995-749A-2  
 ; Sequence 2, Application US/0995749A  
 ; Patent No. US20020155568A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERITDINA HENDRIKA  
 ; APPLICANT: DIJKHUIZEN, LOBBERT  
 ; APPLICANT: RAHMOUI, HAKIM  
 ; APPLICANT: LEER, ROBERT-JAN  
 ; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
 ; FILE REFERENCE: B043388-CIP  
 ; CURRENT APPLICATION NUMBER: US/09/995,749A  
 ; CURRENT FILING DATE: 2001-11-29  
 ; PRIOR APPLICATION NUMBER: 09/604,957  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: EPO 00201871.1  
 ; PRIOR FILING DATE: 2000-05-25  
 ; NUMBER OF SEQ ID NOS: 19

```

; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
; US-09-995-749A.2

```

Query Match 61.8%; Score 68; DB 9; Length 1781;  
 Best Local Similarity 48.0%; Pred. No. 0.055;  
 Matches 12; Conservative 7; Mismatches 2; Indels 4; Gaps 1;

```
QY      1 VPSYSFARAHDS----EVQDIIRDI 21
        :|:||||| |:|:
Db      1123 IPNYSFVARHDNNSQDQIQNAIRDV 1147
```

```

RESULT 7
US-10-417-280A-2
; Sequence 2, Application US/10417280A
; Publication No. US20030229923A1
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: OLANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 0147-0247P
; CURRENT APPLICATION NUMBER: US/10/417,280A
; CURRENT FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: DE 19905069.4
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: US 09/499,203
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: prt
; ORGANISM: Leuconostoc mesenteroides
; US-10-417-280A-2

```

Query Match	61.8%	Score 68	DB 15	Length 2057
Best Local Similarity	63.2%	Pred. No. 0.066		
Matches 12, Conservative	3	Mismatches 4	Indels 0	Gaps 0

```
Qy      1 VPSYSFARAHNSEVQDIIR 19
          :|:|:|:|:|:|:|:|:|
Db      757 IPNYSFVRRAHDYDAQDPPIR 775
```

RESULT 8  
 US-09-955-749A-10  
 Sequence 10, Application US/09995749A  
 Patent No. US20020155681  
 GENERAL INFORMATION:  
 APPLICANT: VAN GEL-SCHOUTEN, GERRITJUDINA HENDRIKARA  
 APPLICANT: DIJKHUIZEN, LUBBERT  
 APPLICANT: RAHAIMI, HAKIM  
 APPLICANT: LEER, ROBERT-JAN  
 TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
 FILE REFERENCE: B043388-CIP  
 CURRENT APPLICATION NUMBER: US/09/955,749A  
 CURRENT FILING DATE: 2001-11-29  
 PRIOR APPLICATION NUMBER: 09/604,957  
 PRIOR FILING DATE: 2000-06-28  
 PRIOR APPLICATION NUMBER: EPO 00201871.1  
 PRIOR FILING DATE: 2000-05-25  
 NUMBER OF SEQ ID NOS: 19  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO. 10  
 LENGTH: 545  
 TYPE: FRT  
 ORGANISM: Streptococcus mutans





```

; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 194106
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(963)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_17303C.1.pap
US-10-424-599-194106

Query Match          42.7%; Score 47; DB 12; Length 963;
Best Local Similarity 42.9%; Pred. No. 81;
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY      2 PSYSPARHDSSEVDIIRDI 22
Db      774 PKLYAMVHDSSVTIPGPFV 794

RESULT 16
US-10-282-122A-71074
; Sequence 71074, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zytkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/220,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

```

; SEQ ID NO 71074
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-71074

Query Match          42.3%; Score 46.5; DB 12; Length 484;
Best Local Similarity 45.5%; Pred. No. 44;
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY      1 VPSYSPARA---HDSSEVDIIR 19
Db      193 VPTYNFAVAVDDHYMQISDVIR 214

RESULT 17
US-10-424-599-267457
; Sequence 267457, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 267457
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(115)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_83537C.1.pap
US-10-424-599-267457

Query Match          41.8%; Score 46; DB 12; Length 115;
Best Local Similarity 55.6%; Pred. No. 9.7;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY      3 SYSFPAHDSSEVDIIR 20
Db      49 STXYALHDSVSTIPGD 66

RESULT 18
US-09-862-027-37
; Sequence 37, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1 Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862,027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345,473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-027-37

Query Match          41.8%; Score 46; DB 9; Length 309;
Best Local Similarity 44.4%; Pred. No. 31;
```

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 5 SPARAHSEVODIIRDI 22  
||:||||:|  
Db 190 SPEKVHDEIKETIGECI 207

RESULT 19  
US-10-276-774-2103  
; Sequence 2103, Application US/10276774  
; Publication No. US20040053245A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y, Tom et al  
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-030  
; CURRENT APPLICATION NUMBER: US/10/276,774  
; PRIOR FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 2700  
; SOFTWARE: Custom  
; SEQ ID NO 2103  
; LENGTH: 490  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-276-774-2103

Query Match 41.8%; Score 46; DB 12; Length 490;  
Best Local Similarity 36.4%; Pred. No. 54;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPQYFARNLPSSKDILOQVV 22  
||:||||:|  
Db 440 VPQYFARNLPSSKDILOQVV 461

RESULT 20  
US-10-131-410-117  
; Sequence 117, Application US/10131410  
; Publication No. US20030235915A1  
; GENERAL INFORMATION:  
; APPLICANT: SPECHT, THOMAS  
; APPLICANT: HINZMANN, BERND  
; APPLICANT: SCHMITT, ARMIN  
; APPLICANT: PILARSKY, CHRISTIAN  
; APPLICANT: DAHL, EDGAR  
; APPLICANT: ROSENTHAL, ANDRE  
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM TISSUE OF BREAST  
; FILE REFERENCE: SCH-1763  
; CURRENT APPLICATION NUMBER: US/10/131,410  
; PRIOR FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: 09/646,673  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: PCT/DE99/00908  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 117  
; LENGTH: 521  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-410-117

Query Match 41.8%; Score 46; DB 15; Length 521;  
Best Local Similarity 36.4%; Pred. No. 58;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPQYFARNLPSSKDILOQVV 22  
||:||||:|

Db 471 VPQYFARNLPSSKDILOQVV 492

RESULT 21  
US-10-131-410-180  
; Sequence 180, Application US/10131410  
; Publication No. US20030235915A1  
; GENERAL INFORMATION:  
; APPLICANT: SPECHT, THOMAS  
; APPLICANT: HINZMANN, BERND  
; APPLICANT: SCHMITT, ARMIN  
; APPLICANT: PILARSKY, CHRISTIAN  
; APPLICANT: DAHL, EDGAR  
; APPLICANT: ROSENTHAL, ANDRE  
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM TISSUE OF BREAST  
; FILE REFERENCE: SCH-1763  
; CURRENT APPLICATION NUMBER: US/10/131,410  
; PRIOR FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: 09/646,673  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: PCT/DE99/00908  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 180  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-410-180

Query Match 41.8%; Score 46; DB 15; Length 564;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPQYFARNLPSSKDILOQVV 22  
||:||||:|  
Db 514 VPQYFARNLPSSKDILOQVV 535

RESULT 22  
US-10-353-929-49  
; Sequence 49, Application US/10353929  
; Publication No. US20030175268A1  
; GENERAL INFORMATION:  
; APPLICANT: ITOH, Kyogo  
; TITLE OF INVENTION: Tumor antigen  
; FILE REFERENCE: GP01-1024  
; CURRENT APPLICATION NUMBER: US/10/353,929  
; PRIOR FILING DATE: 2003-01-30  
; PRIOR APPLICATION NUMBER: JP P2000-231814  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 49  
; LENGTH: 779  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-353-929-49

Query Match 41.8%; Score 46; DB 14; Length 779;  
Best Local Similarity 44.4%; Pred. No. 92;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 5 SPARAHSEVODIIRDI 22  
||:||||:|  
Db 403 SPEKVHDEIKETIGECI 420

RESULT 23  
US-10-433-794-17  
; Sequence 17, Application US/10433794  
; Publication No. US20040077044A1

```
/ GENERAL INFORMATION:
/ APPLICANT: YUE, Henry DING, Li;
/ APPLICANT: IAL, Preeti G.; GRIFFIN, Jennifer A.;
/ APPLICANT: GURURAJAN, Rajagopal, BAUGHN, Mariah R.;
/ APPLICANT: ISON, Craig H.; RAKOWAR, Jayalaxmi;
/ APPLICANT: TRIBOULET, Catherine M.; SWANNAYAR, Anita;
/ APPLICANT: BURFORD, Neil; BANDMAN, Olga;
/ APPLICANT: THORNTON, Michael; KHAN, Farrah A.;
/ APPLICANT: WALIA, Narinder K.; NGUYEN, Daniel B.;
/ APPLICANT: ELLIOTT, Vicki S.; XU, Yuning;
/ APPLICANT: LU, Yan, HARALIA, April J.A.;
/ APPLICANT: VAO, Monique G.; GANDHI, Ameena R.;
/ APPLICANT: ARVIZU, Chandra S.; FORSYTHE, Ian J.
/ TITLE OF INVENTION: KINASES AND PHOSPHATASES
/ FILE REFERENCE: PI-0311 USN
/ CURRENT APPLICATION NUMBER: US/10/433,794
/ CURRENT FILING DATE: 2003-06-04
/ PRIOR APPLICATION NUMBER: PCT/US01/47431
/ PRIOR FILING DATE: 2001-12-04
/ PRIOR APPLICATION NUMBER: US 60/254,034
/ PRIOR FILING DATE: 2000-12-06
/ PRIOR APPLICATION NUMBER: US 60/251,814
/ PRIOR FILING DATE: 2000-12-07
/ PRIOR APPLICATION NUMBER: US 60/255,756
/ PRIOR FILING DATE: 2000-12-14
/ PRIOR APPLICATION NUMBER: US 60/256,172
/ PRIOR FILING DATE: 2000-12-15
/ PRIOR APPLICATION NUMBER: US 60/257,416
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 60/260,912
/ PRIOR FILING DATE: 2001-01-10
/ PRIOR APPLICATION NUMBER: US 60/264,644
/ PRIOR FILING DATE: 2001-01-25
/ PRIOR APPLICATION NUMBER: US 60/266,017
/ PRIOR FILING DATE: 2001-02-02
/ NUMBER OF SEQ ID NOS: 40
/ SOFTWARE: PERL Program
/ SEQ ID NO 17
/ LENGTH: 1345
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No: 7482044CD1
US-10-433-794-17

Query Match          41.8%; Score 46; DB 16; Length 1345;
Best Local Similarity 44.4%; Pred. No. 1.8e+02;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      5 SFARHDSFVQDIIRDI 22
Db      403 SFEKVHDPETKEITGECI 420

RESULT 24
US-10-282-122A-43803
/ Sequence 43803, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zybskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 43803
/ LENGTH: 484
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-282-122A-43803

Query Match          41.4%; Score 45.5; DB 12; Length 484;
Best Local Similarity 40.9%; Pred. No. 64;
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY      1 VPSYSPARA--HDSFVQDIIR 19
Db      193 IFTYPAVALDDHWQISDVIR 214

RESULT 25
US-10-424-599-232822
/ Sequence 232822, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 232822
/ LENGTH: 54
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_52263C.1.pcp
US-10-424-599-232822

Query Match          40.9%; Score 45; DB 12; Length 54;
Best Local Similarity 81.8%; Pred. No. 5.9;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 VPSYSPARA 11
Db      10 VSSYSIARAH 20
```

RESULT 26  
US-10-424-599-229462  
; Sequence 229462, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 229462  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(111)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_49229C.1.pep  
US-10-424-599-229462

Query Match 40.9%; Score 45; DB 12; Length 111;  
Best Local Similarity 44.0%; Pred. No. 14;  
Matches 11; Conservative 4; Mismatches 6; Indels 4; Gaps 1;

QY 1 VPSYSPARAHSEVODI---TRDI 21  
DB 85 LPSYTHRAHATFCADYVDIKRDL 109

RESULT 27  
US-10-424-599-184218  
; Sequence 184218, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 184218  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_137355C.1.pep  
US-10-424-599-184218

Query Match 40.9%; Score 45; DB 12; Length 171;  
Best Local Similarity 42.1%; Pred. No. 23;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVODIIRDI 21  
DB 37 THSFILSDHRSDYEDFLRDI 55

RESULT 28  
US-10-424-599-199858  
; Sequence 199858, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 199858  
; LENGTH: 174  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(174)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_22496C.1.pep  
US-10-424-599-199858

Query Match 40.9%; Score 45; DB 12; Length 174;  
Best Local Similarity 40.0%; Pred. No. 23;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVODIIRDI 22  
DB 141 TVIYAMVHDSVTIPRDFV 160

RESULT 29  
US-09-895-828-454  
; Sequence 454, Application US/09895828  
; Patent No. US20020099012A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Carter, Darlick  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.539  
; CURRENT APPLICATION NUMBER: US/09/895,828  
; CURRENT FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 473  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 454  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_22496C.1.pep  
US-09-895-828-454

Query Match 40.9%; Score 45; DB 9; Length 255;  
Best Local Similarity 38.1%; Pred. No. 36;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSYSPARAHSEVODIIRDI 21  
DB 148 VPSYMSRLTYNQINDVIRKEI 168

RESULT 30  
US-10-114-666-454  
; Sequence 454, Application US/10114666  
; Publication No. US20030103994A1  
; GENERAL INFORMATION:  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY



; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.539C1  
 ; CURRENT APPLICATION NUMBER: US/10/114,666  
 ; CURRENT FILING DATE: 2002-04-01  
 ; NUMBER OF SEQ ID NOS: 479  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 454  
 ; LENGTH: 255  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-114-666-454

Query Match 40.9%; Score 45; DB 14; Length 255;  
 Best Local Similarity 38.1%; Pred. No. 36;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSPARAHDSVQDIIRDI 21  
 Db 148 VPSYKSRRLTYNQINDVIRKEI 168

Search completed: May 4, 2004, 09:25:42  
 Job time : 35.333 secs